DOCUMENT RESUME

ED 132 237 UD 016 650

AUTHOR	Schumer, Florence; Deutsch, Cynthia P.
TITLE	An Exploration of Transfer of Independent Behavior From Enrichment Classrooms in Harlem to a Non-Classroom Behavioral Situation. Final Report for Research Period September 1st, 1969 to August 31st, 1970.
INSTITUTION	New York Univ., N.Y. Inst. for Developmental Studies.
SPONS AGENCY	Office of Economic Opportunity, Washington, D.C.
PUB DATE	Apr 70
CONTRACT	OEO-BCO-5097(A)
NOTE	265p.; For an earlier report, see ED 038 463
EDRS PRICE	MF-$0.83 HC-$14.05 Plus Postage.
DESCRIPTORS	Behavior Patterns; Check Lists; *Enrichment Programs; Grade 3; *Measurement Instruments; *Measurement Techniques; *Minority Group Children; Program Descriptions; Program Evaluation; *Rating Scales; Sociometric Techniques; Student Behavior; Teacher Rating
IDENTIFIERS	*New York (Harlem)

ABSTRACT

The last year of work of a two-year study designed to systematically evaluate the construct of independence in a special ghetto population and its relationship to other variables are described in this report. In this process, several reliable measurement techniques emerge including the classroom behavior checklist, sociometric methods, measures of creativity, and various behavioral situations for assessing independent behavior. The checklist shows excellent potential for use in other investigations, as it yields high independent rater agreement in regard to classifying children into the relevant classifications. Methods for assessing creativity are adopted from existing techniques and again, high interviewer reliability is found. A major result of this work involves offering to the professional community various measurement techniques for further refinement. It is suggested that independence, as a behavioral construct, requires careful definition in the context of the specific exploration in which an investigator is involved. The construct of independence, designated either through teachers' ratings or checklist classifications, does not seem to be related to the kinds of creativity measures and personality variables that are employed. (Author/AM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED132237

FINAL REPORT FOR RESEARCH PERIOD

September 1st, 1969 to August 31st, 1970

OEO Contract No. B00-5097(A)

An Exploration of Transfer of Independent Behavior from Enrichment Classrooms in Harlem to a Non-Classroom Behavioral Situation

Florence Schumer, Ph.D. — Principal Investigator

Cynthia P. Deutsch, Ph.D. — Co-Principal Investigator

Institute for Developmental Studies
School of Education, New York University
Washington Square, New York, N. Y. 10003

Martin Deutsch, Ph.D., Director

UD 016650

April, 1970

DEC 20 1976

Project Staff (1968-1970)

Current Research Coordinator: Peter Pardine, B.A.

Marguerite Levy, Ph.D.

Jo Renée Fine, M.A.

Diana Lackowitz, M.A.

Dorothy Schmidt, M.A.

William Swinburne, B.A.

Patricia Wagner, M.A.

Acknowledgements

We would very much like to thank the children and teachers who have helped us in the execution of our research, and specifically wish to thank the staff and general personnel in the four Harlem schools in which so much of our pilot and experimental activity took place over the two-year research period (1968-1970). These schools are: Public Schools 90, on West 148th Street, 175, on West 134th Street, 79, on East 120th Street, and 68, on West 127th Street. Pilot-testing was done at the Friends Seminary and at Public Schools 121, on East 103rd Street, 80, on East 120th Street, and 57, on East 115th Street. These schools are all in Manhattan, New York City.

In addition to the Lower School Principal of the Friends Seminary, Mrs. Joseph Church, the following people deserve special thanks for aiding us in our work; Mr. Martin Frey and Mr. Edwin Haas, District Superintendents; Mr. Kalman Chase, Dr. Abraham Cohen, Mr. Sol Derewetsky, Mr. Harry Horowitz, Mr. Joseph Mitchell, Mr. John Nailor, and Mr. Leroy Watkins, Principals; Mr. George Carter, Mr. Herbert Coleman, and Mrs. Thomasina Matthews, Assistant Principals; Mrs. Frances Alston, Mr. Kenneth Dawson, Mrs. Rosann De Felice, Mrs. Carol Forster, Mr. Donald Golden, Miss Elna Haynes, Mrs. Holder, Mrs. Odessa Jenkins, Miss Mary Romer, Miss Bonnie Rukin, Miss Lillian Visco, and Mrs. Mary Williams, teachers; and Miss Merle Bullock, Mrs. Annette Jackson, Miss Monica Raymond, Miss Phyllis Sheldon, Mrs. Miriam Weinberg, and Mr. Robert Weinheimer, Assistant Teachers.

We also wish to express our appreciation to members of our own staff for their help in arranging our entry into the schools and for general

guidance in making our way in the schools: Mrs. Jacqueline Stuchin, Curriculum Director, and Miss Edwina Meyers, and Mrs. Laura Schneider Ellis, Curriculum Supervisors. Dr. Carol Silfen, formerly Research Scientist on our staff, also assisted us in connection with some of the behavioral ratings.

Dr. Edith Grotberg and her staff deserve special mention in these acknowledgements in the light of the guidance and advice generously proferred to us during the duration of this study. Finally, special thanks are due to Mrs. Elsie Santiago, Mrs. Lynda Hurwitz, and Miss Carol Kerr for their dedicated secretarial assistance throughout the years.

Contents

		Page
Chapter 1	Background and Hypotheses.....	1
Chapter 2	Related Research.....	9
Chapter 3	Sample and Methodology.....	36
Chapter 4	Scoring Methods, Data Analyses, and Findings.....	60
Chapter 5	Summary of Main Findings and Brief Overview.....	91
References	.....	103
Tables	.....	110
Appendix A	Final Report for Interim Research Period 9/1/68 to 8/31/69	
Appendix B	Creativity Battery: Scoring	

List of Tables

Table

1 Mean Age and Sex for Institute Third-Graders in Public Schools 68 and 79 Classified According to Years of Exposure to the Enrichment Program

2 Mean Age for Subjects Classified According to Years of Exposure to the Enrichment Program for Both Schools Combined

3 Pearson Correlation Coefficients Indexing Inter-Observer Agreement for Behavior Checklist Items

4 Checklist Z-Scores and Ranks for Independence, Dependence, and Pseudo-independence Categories By School

5 Checklist Rank for Each Behavioral Category and Overall Designation for Each Subject

6 Pearson Correlation Coefficients (Interscorer Reliabilities for Two Independent Scorers) for Creativity Measures

7 Test-Retest Reliability of the Sociometric Ratings: Like-Dislike Scale

8 Significance of Differences in Popularity Index (Sociometric Ratings of Like or Dislike) for Children Rated by Own or Other Sex

9 Index of Consistency of Choices (Test-Retest Reliability) in Sociometric Choice Situation

10 Frequency Distributions of Sociometric Choices for Items (1) and (5)

11 Frequency Distribution by Sex of High and Low Scores for the Missouri Children's Picture Series

12 Contingency Data for Teacher's Ratings of S's Reflecting Stability or Change in Ratings

13 Frequency Distribution of Teachers' Ratings of Independence for Fall and Spring Testing Periods in Two Schools

14 Distribution of Two Independent Raters' "Global" Ratings Based on Summary Checklist Data by Category and Degree

15 Relationship Among Checklist Z-Score Designations, Global Checklist Ratings, and Teacher's Ratings (Fall and Spring)

16 Relationship of Checklist Z-Score Classifications and Global Ratings to Sociometric Ratings (Like-Dislike) in the Fall and Spring

17 Relationship of Checklist Z-Score Classifications and Global Ratings to Sociometric Choice Items (1) and (5) in the Fall and Spring

18 Relationship of Sex and Public School to Creativity Scores

19 Relationship of Z-Score Checklist Designations of Independence (I) and Non-Independence (NI) to Creativity Scores

20 Relationship of Global Ratings of Independence (I) and Non-Independence (NI) to Creativity Scores

21 Relationship Between Independence Checklist Measures (Independence vs. Non-Independence) and Missouri Children's Picture Series Subscale Scores (High vs. Low)

22 Relationship of Fall and Spring Teacher Ratings of Independence, Pseudo-independence, and Dependence to Fall and Spring Sociometric Ratings (Like-Dislike)

23 Relationship Between Fall and Spring Teacher Ratings of Independence, Pseudoindependence, and Dependence and Fall and Spring Sociometric Choice Nominations (Items 1 and 5)

24 Relationship of Teachers' Fall Ratings of Independence (I), Dependence and Pseudoindependence (NI) to Creativity Scores

25 Relationship of Teachers' Spring Ratings of Independence (I) and Dependence and Pseudoindependence (NI) to Creativity Scores

26 Relationship Between Teachers' Ratings (Fall and Spring) of Independence vs. Non-Independence and Missouri Childrens' Picture Series Subscale Score (High vs. Low)

27 Frequency Distributions of Global Checklist Ratings, Z-Score Checklist Designations, and Teacher Ratings (Fall and Spring) by Length of Exposure to Institute Program ("Filler Sta us")

28 Relationship of Checklist Z-Score Designations, Global Ratings, and Teacher Ratings of Ss (Fall and Spring) to "Filler" Status

29 Fall Sociometric Rating Ratios (Like-Dislike) by MCPS Designation (High-Low)--(t-Tests)

30 Spring Sociometric Rating Ratios (Like-Dislike) by MCPS Designation (High-Low)--(t-Tests)

31 Relationship Between Fall and Spring Sociometric Choice Nominations (Items 1 and 5) and the Missouri Children's Picture Series Subscale Scores (High vs. Low)

32 Pearson Correlation Coefficients for Sociometric Ratings--Like-Dislike--(Fall and Spring) and Creativity Subtest Scores

33 Point-Biserial Correlations (and t-tests of Significance) Between Sociometric Choice Scores (Items 1 and 5) and Creativity Subtest Scores

34 Mean Creativity Scores for High or Low Status Groups on Missouri Children's Picture Series Subscales

Chapter 1

Background and Hypotheses

Brief Background of Study

In our original study (1968-1969), we attempted to assess the degree to which the Institute's ongoing enrichment program in several Harlem public schools had succeeded in fostering and encouraging independent, autonomous behavior among its pupils. This evaluation was conducted through the use of behavioral rather than standardized test criteria, both in the classroom and in a "transfer" situation, to explore whether or not the changes that occur in the classroom do emerge outside of that programed setting. This investigation and the continuation research (1969-1970)--the latter is described and summarized in the current report, while the former is summarized in Appendix A , in the form of a Final Report for Interim Research Period September 1, 1968, to August 31, 1969--have implications not only for scholastic achievement but also for the child's overall adjustment in daily life. If we have indeed trained for, and reinforced independent behavior in Institute children, we must first be assured that independence should be an educational goal. For this reason, we continued the first year's (initial) research to determine what associated, correlative behaviors and characteristics accompany independent, pseudoindependent, or dependent behavior. The specific objectives of the second year's research were:

(1) To replicate and cross-validate the revised classroom behavior checklist with a new but similar pupil sample. This replication enables us to provide the professional community with a reliable, cross-validated instrument for the measurement of independent, pseudoindependent, and dependent behavior.

(2) To explore the relationship of teachers' ratings of children as to independent, pseudoindependent, and dependent behavior to behavior checklist scores based on observations of these same children.

(3) To explore the sociometric concomitants of independence, pseudoindependence, and dependence.

(4) To explore the relationship of creativity measures and scores to designations of independence, pseudoindependence, and dependence derived from the checklist, and to develop new methods for assessing the creativity dimensions in this sample.

(5) To explore the relationship of certain selected personality variables to designations of independence, pseudoindependence, and dependence derived from the checklist.

(6) To examine a number of additional, subsidiary relationships that can shed light on the behaviors relevant to the area of our investigation.

Our sample was composed of two 1969-1970 third-grade Institute demonstration classes. Third-graders were chosen as these children had the longest exposure to the Institute program and should, therefore, show the greatest effects of this intervention. Within our sample were "fillers," those children who were added to the program because of attrition of the initial group. The presence of "fillers" enabled us to determine whether behavioral change was a function of the length of time the child had spent in the program.

Summary of Procedures

A total of 40 third-grade Institute children from Public Schools 68 and 79 in Harlem comprises the basic sample for the continuation study. These Ss represent the basic subject pool for exploration of the correlates of independence, dependence, and pseudoindependence, and for the cross-validation of the behavioral checklist for independent behavior as described in previous progress reports and in the Final Report for the Interim Research Period (1968-1969).

(1) On the basis of the 1968-1969 study, a revised checklist, consisting of 6 items in the independence scale, 4 items in the dependence scale, and 9 items in the pseudoindependence scale, was cross-validated with the new sample. The items that were used met the criteria of sufficiently high inter-rater reliability and high item-total correlations warranting their continued inclusion in the checklist. The format and procedure established in the 1968-1969 study (see Appendix A) were followed in this cross-validational step, including the use of two independent observers who, early in the school year, recorded by frequency tallies on the checklist, each child's behavior during five-minute periods over a total of 15 sessions on different days.

Data analyses and reliability explorations were handled in a manner similar to that employed in last year's work. Each S was assigned a Z-score for each item of the behavioral categories; his rank of mean Z-scores within each category in terms of the remainder of the sample was determined, and on this basis, he was categorized as independent, pseudoindependent, or dependent. Again, as performed last year, two independent raters globally rated the Ss on the basis of summary sheets of the cross-validated Revised Behavior Checklist; this step was followed by reliability studies of these ratings.

(2) The sociometric measures we developed were administered at the beginning of the school year (Fall, 1969) and again in May, 1970, in accordance with the overall research design of this study. Personnel other than those making the behavioral observations were employed. Reliability examination of all sociometric data will be reported in Chapter 4. The sociometric measures developed by our staff for the current year's work consist of two parts, both given individually. (Pilot-testing had demonstrated the inadequacy of the originally planned group tests.) In part (a) (ratings), each S is asked whether or not he likes each of the children in his class. In the choice test (b), S is asked to select two children for each of the following questions: (a) which two children would you like to sit next to; (b) which two children would you not want to sit next to; (c) which two children would you like to play with during recess; (d) which two children would you not want to play with during recess; (e) name the two children you like best, like very much; and (f) name the two children you don't like at all--they really bug you.

(3) The creativity battery was administered to the above sample in April. This battery consists of four individually given tests, two figural and two verbal, taken from the Torrance Tests of Creative Thinking, Form A (1966). The two figural tests are the Picture Completion test and the Lines test, and the two verbal tests are the Product Improvement test and the Unusual Uses test. Pilot-testing with our population led us to modify slightly the four tests we used in terms of time limits, order of the tests, instructions, and, in one of the subtests, number of stimuli presented to S. Scores were objective, and reliability procedures (scorer agreement) and findings are reported in Chapter 4 .

(4) Completed at the end of April was the administration of the personality and self-image assessment procedure which we have adopted as a

substitute for the Q-sort procedure described in the original proposal. The latter was found to be not feasible in the light of the year's duration of this research. Our thorough investigation of possibilities led us to adopt an objective, easy-to-administer, nonverbal, enjoyable, brief instrument--The Missouri Children's Picture Series, which consists of 238 simple line drawings, each on a 3" x 5" card (Sines, Pauker, & Sines, 1967). Administered individually, the MCPS requires S to place each picture into one of two piles: looks-like-fun or does-not-look-like-fun. Standard administration and scoring procedures for all of the subscales were followed. These procedures permitted the investigators to score for each subscale, as well as develop personality profiles. Scores on this instrument can be transformed to standardized T-scores already established for the MCPS. Split-half and test-retest reliability coefficients are available for this instrument.

(5) Teachers' individual ratings for each child in the above sample were obtained in the Fall. The rating scale was also employed in May, 1970. These scales rate a child's behavior (globally) as independent, dependent, or pseudoindependent. Provisions for considering the intensity of observed behavior (high or moderate) as well as the teacher's confidence in making the rating were included in the rating scales.

Expectations and Hypotheses

Our expectations and hypotheses, expressed in terms of the foregoing objectives and procedures, are:

(1) The checklist will "hold up" in terms of interrater reliability, item-total consistency, and the general distribution of scores that it generated when applied to the new, similar sample (matched in terms of various criteria to the first year's sample).

(2) (a) Teachers' ratings of the children in their classes at the beginning of the school year will show a low (nonsignificant, positive) correlation with scores for the same dimensions (independence, pseudo-independence, dependence) obtained from the behavior checklist; and (b), teachers' ratings of the children in their classes at the end of the school year will show higher (significant, positive) correlations with the same scores derived from the checklist. (Our assumption is that as teachers get to know the children in their classes better, their judgments will become more reliable and valid.)

(3) With regard to the sociometric choices and nominations made by the children about their peers at the beginning of the school year (pretest): (a) such scores will not bear significant relationships to teachers' ratings made at the beginning of the school year; (b) such scores will bear more (but not many) significant relationships to teachers' ratings made at the end of the school year; and (c), such scores will bear significant relationships to checklist scores obtained at the beginning of the school year--specifically, the dependent child will be less likely to be nominated as friend, or as popular, etc., but more likely to be nominated as nonfriend, etc., the pseudoindependent child will be least likely to be nominated as friend, popular, etc., but more likely to be nominated as nonfriend, etc., and that no clearcut relationships will be found with scores for independence.

(4) With regard to the sociometric choices and nominations made by the children about their peers at the end of the school year (posttest): (a) such scores will bear some significant relationships to teachers' ratings made at the end of the school year; (b) such scores will bear

more significant relationships than under (c) above to initial checklist scores, specifically, the dependent child will be even less likely to be nominated as friend, or as popular, etc., but will be even more likely to be nominated as nonfriend, etc., the pseudoindependent child will not be nominated as friend, popular, etc., but will be nominated as nonfriend , etc.; this analysis will also show that independent children will most often be chosen or nominated as friend, popular, etc.

(5) With regard to changes in sociometric status of the children from pre- to posttest: nominations about children designated as dependent and pseudoindependent on the basis of checklist observations will not change significantly, but choices with regard to the children designated as independent will move significantly into the more positive range as friend, popular, etc.

(6) With regard to creativity scores and measures: (a) these will be positively related to designations of independence but negatively related to designations of pseudoindependence and dependence derived from the classroom behavior checklist; (b) these will bear little or no relationships to teachers' ratings of independence, pseudoindependence, and dependence made at the beginning of the school year; and (c) these will bear the same types of relationships described under (a) to teachers' ratings made at the end of the school year, but the extent or magnitude of these relationships will be lower than in (a).

(7) With regard to the personality variables as measured by The Missouri Children's Picture Series, higher scores on subscales prejudged to be representative of "good" or "better" or "healthier" self-concepts or personality variables will be obtained for significantly more children

designated as independent that as pseudoindependent, or dependent on the basis of the classroom checklist; and (b), higher scores on subscales prejudged to be representative of "poor" or "lower" or "unhealthier" self-concepts or personality variables will be obtained for significantly more children designated as pseudoindependent or dependent than as independent on the basis of the classroom checklist. Examples are the Maturity subscale in the former and the Hyperactivity subscale in the latter case. Expectations were not held, it goes without saying, for subscales for which this type of a priori judging was not possible.

(8) With regard to the personality variables as measured by the foregoing instrument: (a) these will bear little or no relationships to teachers' ratings of independence, pseudoindependence, and dependence made at the beginning of the school year; (b) these will bear the same types of relationships described under (a) and (b) under item (7) to teachers' ratings made at the end of the school year but the extent or magnitude of these relationships will be lower than under (a) and (b) under item (7).

Chapter 2 presents an overview of research related to the general area of our investigation, while Chapter 3 describes the sample and methods of our research in considerable detail. The remaining chapters of this report present the findings and discussion of findings of the current exploration.

Chapter 2

Related Research[1]

Our original proposal surveyed some of the literature on independence in children. We will not repeat that review, but instead, after a brief summary of its contents (supplemented by mention of a few additional reports), focus specifically on: (a) researches that involve the use of sociometric methods; (b) researches and reports concerning creativity and its measurement; and (c) material relevant to the personality instrument we have used.

In the development of our own techniques in the currently described investigation, we drew on this literature. Needless to say, reliability and validity considerations as well as methodological feasibility determined the ways in which we used these methods.

Independence in Children--General Survey

As noted in the original proposal, there is a considerable literature on dependence-independence behavior in children, both of an empirical as well as theoretical nature. The reader was referred to Hartup (1963), who presented a comprehensive review of theoretical issues of definition and construct validity, as well as a summary of research in this area. We had also summarized at this time some of the purported correlates of independence, especially with regard to parental and familial behavior, attitudes, and parent-child interaction (Baumrind & Black, 1967; Crandall, Preston, & Rabson, 1960; Finney, 1961; Lederman,

---

[1] This review is taken mainly from the Proposal for Continuation of Research, submitted by the Principal Investigators in April, 1969.

1964; McCord, McCord, & Verden, 1962; Moore, 1965; and Watson, 1957).

Reviewed also were some of the measures and instruments used to assess independent behavior. We noted that these are indeed varied, and include standardized tests, behavioral situations, observations, and rating scales. In the development of the classroom behavior checklist for independence, we reviewed many of these scales and some additional reports by authors such as Linson and Mussen (1957). Especially helpful were reports by Beller (1955), Cervenka (1967), Davidson and Greenberg (1967), Finney (1961), Gewirtz (1956), and Hartup (1963).

Because of its particular relevance to our non-classroom behavioral (transfer) situation, the literature pertaining to behavioral measurement of independence was also reviewed. We found that games, small group situations, play sessions, and the like have been used with various degrees of success. Of most relevance to us was the Asch-type of group conformity situation (1952, 1956). With certain necessary modifications and revisions, we developed Asch-type tasks for our transfer situation. As already noted, the use of classroom observations and ratings as well as nonclassroom behavioral situations had some interesting models from previous research. However, there are few, if any, precedents with regard to the transfer paradigm in explorations with disadvantaged children.

Sociometric Methods

The emergence toward the end of the 1930's of group dynamics as an identifiable field of inquiry was due largely to the influence of Kurt Lewin (1951), whose seminal thinking led to the convergence of separate trends from different disciplines. (See Cartwright & Zander /1960/ for a full discussion of the historical roots and development of the field

of group dynamics.)

One of the many important theoretical and methodological contributions which Lewin made was in demonstrating that complex psychological and social processes could be studied experimentally. He also stressed the importance of the total situation in determining behavior. For Lewin, the total situation included both the physical and social environment as they appeared in the life space of the individual.

An important methodological advance in the study of group behavior was also made by Moreno (1934) through his development of sociometry. During World War I, Moreno had observed that the adjustment of displaced persons seemed to be better when they were allowed to form their own groups within the camp. He later verified this observation by systematic research on groups in institutions (prisons, schools, reformatories) in the United States.

For this research, Moreno constructed a simple questionnaire, on which each person was asked to indicate those other individuals with whom he would prefer to share some specific activity (for example, whom they would like to eat with, work with, live with, sit next to, etc.). These data, reflecting "who chooses whom," can be converted into a "sociogram," or diagram in which the individuals are represented by circles, and the choices represented by lines. The sociogram is, then, a schematic representation of choices and rejections among individuals in the group. There is no single convention for the drawing of these geometric figures, but many alternative ways of representing the relationships. Lindzey and Byrne (1968) have reviewed the alternative methods of representing the data and also the alternative methods of analyzing the data quantitatively.

The most common use of sociometric techniques has been for assessing the attractions and/or repulsions among members of a given group. Moreno advocated the following conditions for using sociometric tests: (a) the limits of the group should be indicated to the subjects; (b) the subjects should be permitted an unlimited number of choices or rejections; (c) the subjects should be asked to indicate the individuals they choose or reject in terms of specific criteria; (d) results of the sociometric questions should be used to restructure the group; (e) the subjects should be permitted to make their choices and rejections privately, without other members of the group being able to identify the responses; and (f) the questions used should be gauged to the level of understanding of the members of the group.

In practice, however, relatively few studies have met all of the criteria which Moreno recommended. Specifically, current studies using the technique of sociometric testing rarely involve the promise of restructuring the group. The technique is also used with groups which are not small enough to be in face-to-face interaction. Furthermore, Moreno assumed as a precondition that the groups under investigation would have existed over an entended time period, while most experimental studies have created ad hoc groups. Another frequent modification has been in the specification of the number of choices the individual is required to make; some investigators have limited the respondents to one, two, or three choices (e.g., most and/or least preferred group members).

Studies using children have frequently supplemented verbal or written inquiries with pictorial devices. For example, McCandless and Marshall (1957) described a "picture sociometric technique" in which each child was asked to point to photographs of playmates preferred for

specific activities. Similar techniques had, of course, been used much earlier by Horowitz (1936), in his "Ranks" and "Show-Me" Tests. The "Show-Me" Test yields social-distance measures. Variations of these techniques have been used in a number of other studies also (for example, Moore & Updegraff, 1964). In a study by Biehler (1954), a "picture completion technique" was used, in which each child was given a line drawing of five figures in a play situation, with instructions to place a photograph of his own face on one of the drawings. Then he was asked to pick the faces of his four favorite playmates to add to the other drawings. Finally, he was given a three-figure drawing and a two-figure drawing with instructions to fill in his favorite playmates. Abel and Sahinkaya (1962), studying the effects of sex and race variables on sociometric choice, presented subjects with pairs of photographs in which the relevant characteristics were systematically varied.

Another technique (first mentioned by Moreno, 1942) which represents an extension of the sociometric test involves the use of sociometric self-rating or prediction of own sociometric rating: for example, "Who do you think will choose you?" Tagiuri (1952) has elaborated this approach as "relational analysis." This technique has been further extended so that subjects are asked to guess <u>all</u> sociometric choices of <u>all</u> members of the groups (Kogan & Tagiuri, 1958; Mann, 1958; Newcomb, 1961, 1963).

The use of "guess who" or identification questions is another technique related to the sociometric test. This method goes back at least to Hartshorne, May, and Maller (1929) who used it in their character studies. In this case, subjects are given various behavior descriptions and asked to guess who among the members of their group is best fitted

by the specific descriptions. This technique is slightly more indirect, so that the intent of the investigator is not so apparent to the subjects. This device, however, is limited to choice or reputation measures, and does not provide information about patterns of interaction or communication.

Some sociometric studies have used the semantic differential method (Osgood, Suci, & Tannenbaum, 1957) for obtaining evaluations of others. By summating evaluative ratings on various scales (for example, good-bad), it is possible to obtain an overall score reflecting attitudes toward another individual (see, for example, Deutsch & Solomon, 1959; and Fishbein, 1965). Petersen, Komorita, and Quag (1964) used a Q-sort technique, whereby subjects sorted, according to their appropriateness for describing specific individuals, a series of sociometric statements into a seven-point distribution.

Lindzey and Byrne (1968) also reviewed research findings from studies involving sociometric status and interpersonal attraction. They reported many investigations of the relationship between adequacy of social adjustment (measured by sociometric techniques) and adequacy of personal adjustment. Bonney (1944) studied 80 children in elementary grades 3 to 5, using the California Test of Personality and a multiple-criterion sociometric questionnaire, and found a correlation of .49 between sociometric status and the total adjustment score of the personality test. Grossman and Wrighter (1948), in a similar study of sixth-grade children, also found that children high in sociometric status scored much higher in total adjustment on the California Test of Personality than those who were low in sociometric status. Lindzey and Byrne (1968) cited studies by many other investigators, using not only the California Test of

Personality, but also the Mooney Problem Check List, teacher ratings, the Minnesota Multiphasic Personality Inventory, the Rorschach test, and the Thematic Apperception Test. The results of these researches suggest a consistent, direct relationship between adequate adjustment and sociometric status.

Several sociometric studies have focused on a specific component of adjustment--anxiety. The Children's Manifest Anxiety Scale has been found to correlate -.30 with sociometric status (Horowitz, 1962; McCandless, Castaneda, & Palermo, 1956; Trent, 1957). The latter author suggested that the negative relationship may be due to the possibility that anxiety distorts the child's perception of others, thus leading to inappropriate interpersonal behavior.

Of direct relevance to the issues under investigation in our own work is Winder and Rau's (1962) work with preadolescent boys (fourth through sixth graders), in which it was found that likability among peers is negatively associated with peer ratings of aggression, dependency, withdrawal, and depression.

In a study by Dunnington (1957b), observers rated preschool children on sociometric status. It was found that children rated high in sociometric status by the observers showed a higher proportion than those rated low in sociometric status of positive affect to negative or aggressive affect, more aggression specific as to origin and object, less verbalization intended to elicit interaction with adults, and more spontaneous verbalization.

Marshall and McCandless (1957) studied the relationship between acceptance by peers and dependence on adults. The sample consisted of 38 children who were attending two newly formed preschool groups. The

observed number of social interactions with adults, while the children were in nursery school free play situations, was used to measure dependency. The authors found an overall negative correlation between dependence and popularity. In addition, the results indicated that sex differences were significant in characterizing the interaction between dependence and popularity. There was a median correlation between dependency and popularity of about -.70 for girls, while for boys the median correlation was about -.40.

With the Marshall and McCandless (1957) study in mind, McCandless, Bilous, and Bennett (1961) explored the relationship between adult-dependency behavior and sex differences in an attempt to replicate and give additional explanations of the original findings of Marshall and McCandless.

Results indicated that emotional dependency and popularity were negatively related. Also, "adult-contacts" and popularity were negatively related. The overall index of adult dependency was found to interfere more with the popularity of girls than with the popularity of boys. However, the difference between sexes was not so great as that in the Marshall and McCandless study.

Using a sociometric interview combining features of both McCandless and Marshall (1957) and Dunnington (1957a), Moore and Updegraff (1964) related sociometric status to sex, age, dependency, and nurturance-giving in three nursery-school groups, ages 3 years, 2 months, to 5 years, 6 months. Results indicated that there was no relation between sex, age, and sociometric status. When sociometric status and nurturance-giving were correlated, the coefficients, although positive, were nonsignificant. However, the results suggested that when a child tends to play a nurturant role in order to satisfy the dependency needs of

others, such a tendency may facilitate, or be facilitated by, his popularity with peers.

In this study, dependency scores were broken down into two components: dependency on children and dependency on adults. When the scores on these two components were related to popularity, results indicated that correlations were all in a positive direction, but were nonsignificant for dependency on children and popularity. In the youngest nursery group, dependency on adults was negatively related to popularity; however, comparisons yielded zero-order correlations in the other two groups. The findings thus suggested that a child's tendency to seek help, reassurance, and attention from other children does not hinder his popularity. It may, in fact, facilitate it. Dependency on adults seemed to interfere with popularity among the youngest children; however, it seems unrelated to popularity among the older nursery school children.

Lindzey and Urdan (1954), in a study using 107 female college student as subjects, found evidence for a curvilinear relationship between sociometric status and measures of dominance and security, indicating that high social status was associated with medium security and dominance.

Fuller and Baune (1951), using third-grade children as subjects, found a significant relationship between sociometric status and accident-proneness. Children high in sociometric choice tended not only to have fewer accidents in general, but also, such accidents as they did have were primarily nonsocial. Children whose sociometric status was low, on the other hand, tended to have a relatively high incidence of accidents which occurred during interaction with other children.

Northway and Wigdor (1947), using Rorschach responses of eighth-grade children stratified into high, low, and middle sociometric status,

found that high-status children showed "greater sensitivity in sensing the feelings of others, and a conscious striving for the approval of others" (p. 196), while the low-status children showed less ability to control their emotions, and seemed to be a more egocentric, moody and impulsive group.

Turner and Vanderlippe (1958) studied the relationship between sociometric status and self-concept, using 175 college students as subjects. Those 25 subjects having the smallest discrepency and those 25 subjects having the largest discrepancy between a Q-sort for self-concept and a Q-sort for ideal self were compared on sociometric status within their residence groups. The investigators found that those with the largest discrepancies were less well liked than those with the smallest discrepancies.

On the other hand, Reese (1961), studying 507 children in the fourth through sixth grades did not find this monotonic relationship. Although Reese also found that subjects with the largest self-ideal discrepancy were the least liked by their peers, children with moderate discrepancies between their self-concept and ideal self were found to be most often choosen by their peers. In other words, Reese found a curvilinear relationship between self-concept and sociometric status, so that children with extremely negative or positive self-concepts were less likely to be chosen than those with moderate self-concepts.

In most studies relating intelligence to sociometric variables, a positive relationship has been found. Bonney (1944), for example, studying elementary school children in the third to fifth grades, found that boys and girls in the upper quartile in sociometric status were also distinctly superior to those in the lower quartile in sociometric status

as measured by group intelligence tests. Grossman and Wrighter's (1948) findings on sixth-graders also showed a positive relationship between sociometric status and Stanford-Binet scores, as well as between sociometric status and reading ability. (See Lindzey and Byrne [1968] for additional studies relating intelligence to sociometric variables.)

Evidence that within-sex sociometric choices exceed cross-sex choices has been reported by Moreno (1934), and many later investigators have substantiated this finding across children of all ages from nursery school through the teens. Nevertheless, Kanous, Daugherty, and Cohn (1962) have suggested that this preference-for-own-sex phenomenon may apply only to middle-class subjects. Kanous, et al. studied 375 children in the second through eighth grades, and found that when these children were divided into upper and lower socioeconomic strata, those in the lower socioeconomic level were significantly less likely than those in the upper socioeconomic level to limit choices to their own sex.

Creativity

The literature on creativity can be characterized by lack of consensus as to its definition, empirical correlates, and techniques for its measurement. Yet, most investigators and workers in the general area of child development would agree that the construct "creative" can be meaningfully applied to some individuals and not to others. Reports in this area are based on different interpretations as to the meaning of creativity; and subject selection--that is, how the creative Ss and the noncreative Ss are to be chosen -is based on varying types of criteria. For example, creativity has been defined as the cognitive quality measured by specific paper-and-pencil tests (Getzels & Jackson, 1962); as a thought process involving the sensing of, and dealing with,

difficulties, problems, and gaps (Torrance, 1965); and the quality and quantity of ideas produced by members of a professional group (Meer & Stein, 1955). While a majority of the investigations to be discussed have employed samples of creative individuals selected on the basis of paper-and-pencil tests of "creativity," some investigators have used behavioral criteria. Other investigators have defined a creative person as one who has already demonstrated creative ability (Butcher, 1967). Generally, samples of creative adults are chosen from populations of individuals who have produced creative products, for example, artists or scientists, while samples of creative children have been drawn on the basis of paper-and-pencil tests or ratings and judgments by peers or teachers. The problem thus arises as to whether findings for adults can be generalized to children and vice versa, for not only are sample selection procedures different, but also, a general question can be raised as to whether a child who has creative potential as indicated by a creativity measure will eventually demonstrate creative performance later on in life.

Behavioral tasks. As noted, we have chosen to develop behavioral measures for assessing our Ss' creativity. For this purpose, we surveyed available models in the literature to see if they could be adopted for our own uses. This section describes the possibilities that we considered.

The basic test of creative thinking abilities utilized by Torrance (1965) in most of his studies is the Test of Imagination (Form D). The five tasks on this test, Product Improvement, Product Utilization, Unusual Uses of Tin Cans, Circles, and Ask and Guess, are briefly described below: (a) Product Improvement--subjects are shown a toy dog and asked

to list the most interesting, clever, and unusual ways for changing the toy so that children would have more fun playing with it; (b) Product Utilization--subjects are asked to list the most interesting, clever, and unusual uses they can think of for the toy dog other than as a plaything; (c) Unusual Uses of Tin Cans--subjects are asked to list the most unusual, interesting, and clever ideas they can think of for using tin cans; (d) Circles--subjects are asked to see how many objects they can sketch which have a circle as the main part; and (e) Ask and Guess--subjects are shown a picture (Torrance used a picture of Tom, the Piper's Son) and are asked to write down all the questions they can think of about things they see in the picture.

Two forms of these tests were designed as alternate-form tests for kindergarten through third grade, and two additional forms were designed as alternate forms for the fourth, fifth, and sixth grades.

The following are tests used by Getzels and Jackson (1962) to measure creativity: (a) Word Associations--subjects are presented with 25 words that have multiple meanings, for example, arm or punch, and are asked to write as many meanings as possible; (b) Uses Test--subjects are given five common objects, for example, bricks or pencils, and are asked to think of as many uses for them as possible; (c) Hidden Shapes--subjects are shown eighteen simple geometric figures and four complex figures and are asked to identify the complex figure in which the simple figure appears; (d) Fables--subjects are given four fables whose last lines are missing, and are asked to supply a moralistic, humorous, and a sad ending; and (e) Make-Up Problems--subjects are given four complex paragraphs, each containing many numerical statements about activities such as buying a house, or building a swimming pool, and are then asked

to make up as many numerical problems as possible.

Paper-and-pencil tests; peer and/or superiors' ratings. Questions concerning the validity of paper-and-pencil tests of creativity seem to arise most frequently in the investigations of the relation of creativity and IQ. To illustrate: Getzels and Jackson (1962) stated that no significant correlation was found between creativity tests and IQ. Wallach and Kogan (1965), however, noted that the five tests of creativity used by Getzels and Jackson were no more strongly correlated with one another than they were with intelligence.

Vernon (1967) concluded that the question of whether divergent thinking abilities ("thought that issues in a variety of new answers" /p. 153/) are part of, or distinct from, what is measured by intelligence tests depends more on the form of the tests and the way they are given and scored than on their apparent creative or noncreative content.

Ward (1968) reported that he found the same high correlations among Wallach and Kogan's measures with seven- and eight-year-old boys that the latter authors had found for older children. He added, however, that the findings for these measures with a kindergarten sample were ambiguous, and suggested that there may be no unitary creativity dimension for children younger than elementary school age or that possibly the particular measures in question might not be appropriate for this age group.

Rather than administering paper-and-pencil tests of creativity, some investigators have preferred to assess their Ss' creativity by having them rated by peers (see, for example, Reid, King, & Wickwire, 1959), or by their superiors--teachers, department heads, etc. (see Crutchfield, 1962; Meer & Stein, 1955; and Stein, 1963).

Yamamoto, Lembright, and Corrigan (1966) found that paper-and-pencil

measures of creative thinking were not highly and consistently correlated with peer nominations or comparable creativity measures among high school students. They hypothesized that measures of perceived degree or level of creativity (based on peer nominations on creativity) were better predictors than objective measures.

Wallach and Kogan (1965), however, stated that, "given the dubious state of creativity ratings in the case of members of professional groups for whom high level products are available, teacher and peer ratings of creativity must by comparison be evaluated as next to worthless" (p. 68-69). That teacher ratings of creativity can be potentially valid indicators of this construct was suggested by Torrance (1965) who stated that children nominated by their teachers on creativity criteria achieved higher scores on tests of creative thinking than did their peers not so nominated or than did their peers who were nominated as being especially low on the criterion in question. Torrance noted that in peer nomination studies, results have been significant above the third-grade level--that is, above this level, pupils receiving a large number of peer nominations on various criteria of creative thinking achieved higher scores on the corresponding measures of creative thinking.

Creativity and sociometric status. Findings concerning the relationship between measured creativity and sociometric status are by no means definitive. If any general summary can be made, it would be that the child who is highly intelligent as well as highly creative will be most frequently chosen by his/her peers. Even this statement, however, may not be true for all grade levels or for both sexes. Furthermore, all relevant reports used middle-class white samples. Only one investigation was based on children in the early elementary school grades

(Torrance, 1963), and no studies were conducted with preschool children.

In Torrance's study (op. cit.), sociometric nominations were obtained in thirty-three Minneapolis public school classrooms (grades one through six). Children were asked to name: (a) their best friend; (b) those who talk the most; (c) those who have the most good ideas; (d) those who think of the most ideas for being naughty; (e) those who think of the most wild or silly ideas; and (f) those who hold back their ideas.

Some of Torrance's findings were: (a) in the first and second grades, there was a general tendency for the most highly chosen child on the good-ideas criterion to be only moderately high on measures of creativity; there was no other consistent pattern except that highly creative boys tended to be nominated on naughty or silly ideas; (b) in the third grade, high-creative girls were chosen on the good-ideas criterion, but this was not found for high-creative boys who tended to be highly chosen for naughty and silly ideas; (c) in the fourth grade, very few of the highly chosen children on good ideas fell into the high-creative category, but high-creative children did receive a moderate number of nominations on this criterion; (d) in the fifth grade, high-creative children were nominated fairly frequently on the good-ideas criterion; and (e) in the sixth grade, most of the children who scored high on the tests of creativity were chosen for having good ideas--a finding which was statistically highly significant. Other findings were that high IQ children who were lower on creativity tended to have the most friends, while those who were high in creativity but lower in IQ had the fewest. Those children who were high on both creativity and IQ received the most nominations for talkativeness, good ideas, naughty ideas, and silly ideas. The author's general conclusion was that apparently classroom groups do not

perceive their most creative members as having good ideas until the fifth or sixth grades. (Note: when Torrance and others define a group as "low IQ," it should be understood that "low" is in relation to the rest of their sample; since their Ss tend to be average or above in IQ, these data should be regarded with some caution in terms of its application to low SES groups with lower mean IQs.)

In an additional study, Torrance (1963) studied twenty-five groups, taking one class at each grade level from two to six. He found clear evidence of peer pressure against the most creative member in all 25 groups. Observers noted that 68% of the most creative children initiated more ideas than any other member but only 24% of these children were seen by their peers as making the most valuable contribution to group performance.

Yamamoto studied creativity and sociometric choice among adolescents (1964) and fifth graders (Yamamoto, Lembright, & Corrigan, 1966). In the former study, Yamamoto studied the peer nominations of 428 University of Minnesota high-school students in grades seven to twelve. He found that the high, middle, and low creativity groups tended to receive peer nominations for creativity according to expectation (high>middle>low) in junior high school grades; but in senior high school grades, this was the case only among the boys. On the best-friends criterion, it was found that highly creative adolescents were more popular among boys than among girls, but that highly creative adolescents, as a group, received no more peer nominations than did the less creative adolescents.

In the latter investigation, 730 fifth-grade white pupils from a middle-sized industrial city suburb were studied. For these children, there were significant but low, positive correlations among nominations

as friends, creative-thinking scores, and intelligence scores, while correlations among number of nonfriends nominations and the two variables were low, significant, and negative.

The authors concluded that creative thinking as well as intelligence did not carry much weight in explaining a child's sociometric status in the classroom. They suggested, however, that the more intelligent and creative a child, the more likely he would be nominated as a friend and less often as a nonfriend.

Kurtzman (1967), in a study of 80 male and 71 female ninth-graders in which intelligence level was controlled, found that the high-creative girls were less well accepted by their peers than either the middle- or low-creative girls. High-creative boys, however, were the most well accepted by their peers.

Wallach and Kogan (1965) studied 70 male and 81 female middle-class fifth graders who were divided into four groups: high creative, high IQ; high creative, low IQ; low creative, high IQ; low creative, low IQ. Based on ratings by observers, these authors found no significant relationships among creativity, IQ, and peer popularity for boys. Findings for girls, however, were significant, with high-creative, high IQ girls most sought after, and most seeking others, while high-creative, low IQ girls were neither sought after nor seeking of others. Low-creative high IQ girls were sought after, but did not seek others, while low-creative, low IQ girls were not sought after, but did seek others. The high-creative low IQ girls were the most isolated in the classroom, and were found to be withdrawn, lacking in self-confidence, inclined toward disruptive behavior, and attention-seeking.

In an illustrative study that did not involve specific sociometric

ratings, Holland (1961) measured the relationship of creative achievement and self-ratings of popularity in mailed questionnaires to 994 high school juniors who were National-Merit finalists. For girls, no correlation was found between self-ratings of popularity and artistic or scientific achievement. For boys, there was no correlation between popularity and scientific performance, but there was a significant relationship between popularity and artistic creativity.

Creativity and independence. Although there are no major studies dealing with the relationship of creativity to independence in children, many subsidiary findings of a number of investigations suggest that creative children and adults exhibit more independence in behavior and thinking than do their non-creative peers.

Kurtzman (1967), in a ninth-grade sample, found a consistent trend, although not always statistically significant for both sexes, for high-creative children to be considered more "adventurous." Torrance (1963) noted that high-creative and high IQ children were rated by their teachers as more independent. And Holland (1961), found that male National-Merit finalists with high scores on a creative-science and creative-arts scale, rated themselves as independent and received high scores on independence of judgment. For girls, the same relationship was found with self-ratings of independence, but not with scores on independence of judgment.

Getzels and Jackson (1962) observed in their sixth- to twelfth-grade sample a "certain mocking attitude on the part of the creatives toward conformity and toward conventional success" (p. 41). The authors suggested that the independence of the highly creative children may be a factor in determining their friendship nominations by peers. They

noted that "the creative students may indeed be more difficult to get along with. But it must be realized that their behavior may have its source in independence of thought rather than in malice" (p. 125).

The following studies have attempted to explore the relationship of independence to creativity in adults. Butcher (1967) compared 63 visual artists and 28 craft students to matched controls. The artists and students were significantly more assertive, dominant, and self-sufficient than the controls. Industrial research chemists designated as creative by their superiors, colleagues, and subordinates were studied by Stein (1963). These scientists were found to have a high degree of automony, self-sufficiency, and self-direction. They exhibited a marked independence of judgment and a rejection of group pressures toward conformity thinking. This finding was supported by Barron (1953), who studied iniividuals who refused to yield to strong pressure from their peers to concur in a false group opinion. The non-yielders, as contrasted with the yielders, described themselves more often as "original" and "artistic."

Crutchfield (1962) conducted three investigations of the relationship between creativity and conformity. In the first study, 45 scientists rated highest in originality by an assessment staff were found to be less conforming than those rated lower. In the second investigation, involving 24 seniors at Mills College, it was found that the 12 rated as outstanding in creative ability by the faculty received markedly lower scores on a conformity test. Finally, 40 architects who were nominated as being among the most creative in their profession were compared to 84 architects not considered to be creative. The latter group scored significantly higher on a conformity scale than the more creative group.

Creativity, achievement, and intelligence. Findings regarding the relationship of creativity to achievement are inconclusive, and may well remain so until the interactive effects of intelligence and creativity on achievement are better understood. Some aspects of the creativity-IQ issue have been briefly mentioned earlier. It should be reiterated, however, that there is some difference of opinion among investigators as to the correlation of these two variables.

Most authors seem to accept the distinction between IQ and creativity. Getzels and Jackson (1962) and Torrance (1965) reported that the correlation between intelligence and creativity was very low, and Ward (1968) found nonsignificant relationships between IQ and creativity measures for 34 seven- and eight-year-old boys.

Madaus (1967), in a study of over 600 ninth- and tenth-grade students, found a negligible relationship between measures of intelligence (school and college-ability test) and the Minnesota Tests of Creative Thinking. Palm (1959) and Bentley (1961) also found low correlation coefficients between a battery of the Minnesota Tests of Creative Thinking and the Miller's Analogies Test.

Contrasting results were described by Iscoe and Pierce-Jones (1964) and Meer and Stein (1955). The former authors, testing 220 lower SES Texas children, half Negro and half white, found a low but statistically significant correlation between tests of divergent thinking and IQ. Meer and Stein, using a sample of 64 research chemists ranked in respect to creativity by their department heads, found that for all subjects there was a significant relation between intelligence and creativity. When education was held constant, however, there was a significant relationship only among the non-Ph.D.s between creativity and intelligence.

In addition to negating any relationship between creativity and intelligence, Getzels and Jackson (1962) and Torrance (1960) found that creativity is as closely related to school achievement as is intelligence. The former investigators noted that the correlation between certain of the creativity tests and school achievement for their whole school population (grades six to twelve) equalled, or in some cases, exceeded, the correlation between IQ and school achievement. The correlations between verbal achievement and three of the creativity tests were as great or greater than between verbal achievement and IQ. Torrance, replicating these studies, found similar results in six of eight schools studied. Both the high IQ and the high creativity groups were equally superior in achievement to the population from which they were drawn.

Karsten (1968) administered six of Guilford's tests of creativity to 201 thirteen-year-old Negro pupils. She found that the measure of achievement, grade-point-average, was independent of two of the measures: "seeing problems," and "utility." Grade-point-average for both advantaged and disadvantaged Negro youth was significantly related, however, to other creativity and intelligence measures. Significant sex differences in favor of girls in grade-point-average were also found.

Twenty-four creative and twenty-four noncreative seventh graders (defined by peer nominations) were compared by Reid, King, and Wickwire (1959). The creative groups exhibited superior achievement on almost all cognitive variables. However, the authors found that cognitive variables, as measured by general intelligence and achievement instruments, were closely related to the perception of children as creative by their peers. Further findings indicated that on reading comprehension and the mechanics of English, performance by upper-status girls (based on

an SES scale) was uniformly high regardless of creativity grouping, and that lower-status boys scored uniformly low regardless of creativity grouping. Scores of creative upper-status boys and of creative lower-status girls were significantly superior to their respective noncreative counterparts.

Eiseman, Platt, and Darbes (1968) administered two tests of creativity, one test of intelligence, and one achievement test to 62 high school students. Six 2 x 2 analyses were computed for each achievement subtest, with high and low creativity and high and low IQ as independent variables. IQ was significantly related to five of six subtests (paragraph meaning, word meaning, spelling, language, and arithmetic composition). Creativity was significantly related only to scores on arithmetic reasoning. The investigators further found that there was no significant IQ x creativity interaction, although the subjects who were low on both variables had the lowest scores on all six subtests.

Laughlin, Doherty, and Dunn (1968) and Laughlin (1967) found that among high school and college students, creativity was significantly related to success on an incidental concept-formation task. It is suggested that creativity may be a more sensitive prediction for incidental learning than for intentional learning, and that the ability to form, retain, and utilize remote associations may be an underlying process in both creativity and incidental learning.

The findings of the above studies, that achievement and creativity are positively related, are contraindicated by other investigations. Holland (1961), in his study of creative and academic performance among National-Merit finalists, concluded that for samples of students of superior scholastic aptitude, creative performance is generally unrelated

to scholastic achievement and scholastic aptitude.

Flescher (1963) stated that he found no evidence for the conclusion that creativity is as closely related to achievement as is IQ, and attributes this to: (a) a high correlation between many creativity and IQ tests; (b) an extremely low correlation between a creativity index and IQ; and (c) the questionable validity of creativity tests.

Edwards and Tyler (1965) compared 181 ninth-graders on scores for two of the Minnesota Tests of Creativity, the School and College Ability Tests (SCAT), and the Sequential Tests of Educational Progress (STEP). The high SCAT, low creativity group was found to be significantly superior to the low SCAT, high creativity group on both grade-point-average and the STEP. The same authors also tested Torrance's threshold concept that below some critical point (about 115 to 120 IQ), differences in IQ are the major determinants of differences in academic achievement, and that above this point differences in achievement are thought to be more closely related to differences in creativity. These investigators concluded that the "threshold concept" was not in operation in their sample.

Finally, Hasan and Butcher (1967), replicating the study of Getzels and Jackson (1962), found that in a sample of twelve-year-old Scottish boys, the high-creative Ss did not obtain as high attainment scores as the high IQ Ss. Taft (1967), studying 107 fourteen- to seventeen-year-old Australian boys, found that creativity tests had little correlation with the Junior Exam (a measure of school achievement).

The Missouri Children's Picture Series: Assessment of Personality and Self-Concept

Because we were interested in studying self-concepts in relationship to the independence dimension in our sample, we had explored various possibilities for self-concept assessment in our particular population. We had decided, after some consideration, to explore the possibility of developing a technique based on Q-sort methodology, and presented a detailed account of relevant literature (concerning objectivity, reliability, and ease of administration) in the Proposal for Continuation of Research, submitted in April, 1969. We introduced that section by describing some methodological considerations concerning the measurement of self-concept, and indicated that Bennett (1964), Cronbach and Meehl (1955), Crowne and Stephens (1961), and Payne and Farquhar (1962) have all considered several problems in operationally defining self-concept.

Bennett (op. cit.), in a review of several studies, reported that existing techniques for measuring self-concepts were vulnerable to a number of criticisms. Among these are the subjectivity of the checklist, the difficulty of consistency in scoring an answer to an open-ended question, and the arbitrary labeling of clusters derived from factor-analytic methods.

In the proposal for continuation research noted above, it was observed that only one study had been reported in which the Q-sort instrument was employed with a young, elementary school (third-grade) sample (Bennett, 1964). As described in the literature (Nunnally, 1959; Stephenson, 1953; Wittenborn, 1961) this technique has been widely and successfully used with older student populations (Quimby, 1967) and adult populations (Medinnus, 1961; Rogers & Dymond, 1954; Walker, 1968).

In the present attempt to explore the possibility of employing Q-sort methodology with disadvantaged third graders, various problems arose which precluded the development of a reliable instrument. In our pilot-phase experiences, for example, we observed that children at this grade level do not readily understand the instructions required for performance. This appeared specifically to involve the children's difficulty in clearly understanding fundamental relationships required for performance. Additional problems involving reading levels as well as the pure mechanics of administration of this particular instrument required too expensive a pilot investigation in the light of the time limits of the study.

Because of our major concern with item-difficulty as related to reading comprehension levels of the subjects, an instrument, The Missouri Children's Picture Series (Sines, Pauker, & Sines, 1967) seemed relevant to our exploratory aims. Our pilot investigation of various possibilities led us to adopt this instrument which is an objective, nonverbal, enjoyable test situation designed for easy administration and scoring. It consists of 238 simple line drawings, each on a 3" by 5" card. Administered individually, the MCPS requires the subject to place each picture into one of two piles: looks-like-fun or does-not-look-like-fun. The standard administration and scoring procedures presented by the test-developers were followed, permitting the investigators to score for each subscale of the instrument.

The MCPS has been examined quite rigorously for reliability and validity purposes (Sines, Pauker, & Sines, 1968). For example, each subscale has been examined for internal consistency and stability. Furthermore, extensive research has been carried out exploring the validity of

the instrument (Baker, 1969; L'Abate & Hosford, 1967; Owen, 1968; and Sines, Pauker, Sines & Owen, 1969). Criterion measures have ranged from checklist data supplied by parents to institutional and clinically relevant behavior dimensions. The following subscales of the instrument have been scored and employed in current data analyses: Maturity, Conformity, Inhibition, Aggressivity, and Hyperactivity. The scores derived from each subscale are in the form of a standardized T-score, established by sex and age of the subject.

Chapter 3

Sample and Methodology

Chapter 1 presents the general model for the currently described research (1969-1970), while Appendix A presents the completed presentation of method and findings for the preceding year's work. The second year's work included: (a) the cross-validation (Fall, 1969) of the Revised Classroom Behavior Checklist, which employed two independent observers who recorded by frequency tallies on the checklist each child's behavior during five-minute periods over a total of 15 sessions on different days; (b) data analyses and reliability explorations of the material obtained under (a) which included the development of Z-scores within each category (independent, pseudoindependent, and dependent), the assignment of the Ss into an appropriate behavioral category depending on his rank of mean Z-scores within each category, the exploration of item-total correlations within each category, the computation of observer-reliability of each item, the global rating of each S by two independent raters on the basis of summary sheets of the cross validated Revised Behavior Checklist, and the reliability analysis of these global ratings; (c) the development and reliability analysis of sociometric classroom measures which were individually administered to the Ss at the beginning of the school year (Fall, 1969) and again in May, 1970, in accordance with the overall research design of this study; (d) the development and reliability analysis of a creativity battery (actually adapted from the Torrance Tests of Creative Thinking, Form A (1966)) which was individually administered to the Ss in April, 1970; (e) the individual administration of The Missouri Children's Picture Series (Sines, Pauker, & Sines, 1967) to all of the Ss, a procedure which was completed by the

end of April, 1970; and (f) the obtaining of individual teachers' ratings for each S in our sample along a discontinuous scale for independence, pseudoindependence, and dependence in the Fall of 1969, and again in May, 1970.

The current chapter describes these procedures and the subject population we employed in their development and use.

Sample

Because the currently described study was of one year's duration, we could not work with all of the third graders in the Institute's demonstration classes in three Public Schools in Harlem. To obtain a feasible sample size, two of the schools were selected on the basis of their yielding the largest number of Institute third-grade children who have had at least two years of exposure to our enrichment program. These schools were Public Schools 68 and 79 in Harlem.

(1) Basic Sample--Classroom Observations. The distribution of this sample by school, age, sex, and length of time in the Institute program is shown in Table 1. As can be seen in this table, 18 children were observed at Public School 68 and 22 children at Public School 79. Of the 40 children included in this sample, 19 had at least two years of exposure to the Institute's program (10 at Public School 68 and 9 at Public School 79). Table 1 shows that average age for each school is virtually the same, and that only minor age differences for the various types of Ss within each of the schools are found. Most Ss, as of September, 1969, were eight years and a few months old. Examination of this table also indicates that the sample composition is fairly evenly distributed by sex.

Since years of exposure to the IDS program is a variable in which we are interested, and since the development of independence is

age-related, it is interesting to see whether or not wide differences in mean ages are found between subjects classified as to "filler" status. Table 2, which presents mean age for various classifications of Ss (with both schools combined), indicates that as of September, 1969, average age for each classification is approximately 8 years and a few months; wide differences between groups on this variable did not emerge.

There was a slight degree of sample attrition during the academic year, because of the transfer of one subject out of Public School 68 after the checklist observations were completed. Consequently, the possible N for each of the instrument administrations was 39, with the exception of the classroom observations (and, of course, global checklist rating) and teachers' Fall ratings, where N was 40. Following is a description of the number of subjects for each of the quantitative measures used in this study.

(2) Creativity Measures. For the Verbal battery, for the above reason (school transfer), N was 39, but for the Figural battery, N was 38, because one child was not testable.

(3) Sociometric Measures. N was 39.

(4) The Missouri Children's Picture Series. N was 39. However, when findings based on statistical analysis are presented, it will be seen that smaller numbers of Ss were used for purposes of the particular analyses involved, since cases whose rank fell at the median of the distribution of scores were eliminated. The following N's resulted: Conformity: N of 31; Maturity: N of 37; Aggressivity: N of 30; Inhibition: N of 38; and Hyperactivity: N of 31.

(5) Teachers' Ratings. In the Fall, N was 40, and in the Spring, N was 39, for reasons already noted. As will be seen, however, for purposes of analysis, actual N's that were employed were 38 and 36

respectively, since two "Can't Rates" were employed in the Fall ratings, and three "Can't Rates" in the Spring.

(6) Global Checklist Ratings. N was 40. For purposes of analysis, of course, fewer cases were employed; since 18 of the ratings represented disagreements, only 22 Ss were used for correlation purposes with other measures.

Methods

A. The Revised Classroom Behavior Checklist

Working definitions of behavioral categories and development of the Classroom Behavior Checklist. The Final Report for the Interim Research period (see Appendix A) carefully outlined pilot and test phase activities involved in the development and eventual revision of the behavioral classroom checklist in 1968-1969. The working definitions of the behavioral categories are reflected below.

(1) Independent behavior is oriented toward productive (constructive) goals, is age-related, and is self-initiated. The independent child manifests behavior which is both self-reliant and self-assertive. He is considered independent of others to the extent that he attempts to satisfy his needs in a manner which is not entirely a function of the way others behave toward him. A child is considered independent when he conducts his problem-solving activities without seeking unnecessary help from others, and is able to carry on his activities without continuous need for reassurance, affection, or approval. Independence also includes behavior which is self-assertive, such as striving to master tasks. Independent children characteristically show relatively good motor control; that is, their mobility is appropriate to the situation: they are neither overly constricted nor excessively motoric.

(2) Pseudoindependent behavior can be described as nonproductive, nondependent behavior, characterized by the seeking of attention through asocial or antisocial activities, the maintenance of physical and psychological distance from peers or teachers (leaders), high motility, and inconsistency in the pursuit of goals. Children so designated are likely to interfere with the activities of other children, refuse to participate in group activities, and be destructive of work materials and equipment. Such children often do not accept help or follow instructions, even though their own efforts are clearly not successful. They also tend to show random sequences of behavior with unrelated activities begun and discarded before completion.

(3) Dependent behavior is determined by the presence of specific characteristics, not merely by the absence of characteristics specific to independence. Responses that are characterized as dependent are often those that seem directed toward obtaining social reinforcement. Such behavior includes many different kinds of activity, such as: persistently seeking help (whether necessary or unnecessary); affection-seeking; and bids for attention, reassurance, approval, succorance, and guidance. Dependent behavior is also manifested in passivity and lack of spontaneity.

The checklist resulting from these definitions was studied intensively both in its observational phase and its analysis phase. Items remaining for the Revised Behavior Checklist, for example, were those items which met the criteria of sufficiently high inter-rater reliability and high item-total correlation coefficients to warrant their continued inclusion in the final pool. This revised checklist appears below.

CODE _____ SCHOOL _____ TEACHER _____ SEX _____ OBSERVER _____ DATE _____ STARTED _____ NUMBER _____

The Revised Classroom Behavior Checklist

	Frequently	Moderately Often	Rarely	Never	No Opportunity to Observe
(1) withdraws in the face of difficulty					
(2) is inattentive; is easily distracted by things going on around him					
(3) uses materials and equipment in a rough or destructive manner					
(4) helps, sympathizes, shows consideration of others; is thoughtful; offers praise, guidance, assurance					
(5) sits quietly and attentively when task requires					
(6) brags, shows off, displays exaggerated opinion of own abilities					
(7) seeks attention, i.e., in a manner that seems designed for active notice rather than stemming from a need for help					
(8) utilizes available free time constructively					
(9) interferes with the work of others					
(10) withstands interference while engaged in individual activities, e.g., doing difficult assignment, a puzzle, painting					
(11) does not follow directions					
(12) requires close assistance or direction in order to work at a task					
(13) seeks to do things differently from others, even when own method is not effective					
(14) lets other children impose on him or boss him around					
(15) tries to figure out things for himself before calling on teacher (leader) or other children for help					
(16) does not pursue task to completion					
(17) seeks constant recognition and/or approval					
(18) speaks out of turn; interrupts others; takes other's turn, etc.					
(19) attempts to resolve difficulties that arise between himself and other children without appealing to teacher (leader)					

The foregoing behavior checklist was employed by two independent observers of all the children in the same classroom situation (at different times during at least a two-hour period at each school) on different days over a period of almost two months. The sequence of items is a random one. Below are the checklist items (with examples of actual behavior) for the Revised Checklist regrouped into categories according to what we think they are reflecting, as judged by construct considerations as well as empirical considerations during pilot and test phases of our exploration.

Independence Items--6

(4) Helps, sympathizes, shows consideration of others, is thoughtful, offers praise, guidance, assurance--e.g., helps others with work when appropriate, lends or shows materials freely.

(5) Sits quietly and attentively when task requires--e.g., remains undistracted by irrelevant activities.

(8) Utilizes available free time constructively--e.g., initiates productive activity when there is no ongoing class activity.

(10) Withstands interference while engaged in individual activities, as in doing difficult assignment, puzzle, painting, etc.--e.g., does not abandon task when interrupted.

(15) Tries to figure out things for himself before calling on teacher (leader) or other children for help--e.g., tries to overcome obstacles in the environment.

(19) Attempts to resolve difficulties that arise between himself and other children without appealing to teacher (leader)--e.g., does not initially seek adult intervention to solve difficulties between himself and peers.

Dependence Items--4

(1) Withdraws in the face of difficulty--e.g., abandons goal when presented with an obstacle.

(12) Requires close assistance or direction in order to work at a task--e.g., easily distracted; must have teacher or assistant's help to remain attentive to task.

(14) Lets other children impose on him or boss him around--e.g., does not stand up for his rights; yields his place in line, relinquishes turn at play, etc.; allows others to exploit him.

(17) Seeks constant recognition and/or approval--e.g., tries to get teacher's or assistant's attention as frequently as he can.

Pseudoindependence Items--9

(2) Is inattentive, is easily distracted by things going on around him--e.g., engages in random class-unrelated activity (wanders about, daydreams when he should be working on own); talks to visitors about irrelevant matters when he/she should be engaged in a task; daydreams; attention wanders from task at hand; is not prepared to answer when called upon in class.

(3) Uses materials and equipment in a rough or destructive manner--e.g., destroys books, crayons, games, etc.

(6) Brags, shows off, displays exaggerated opinion of own abilities.

(7) Seeks attention, i.e., in a manner that seems designed for active notice rather than stemming from a need for help--e.g., is inappropriately aggressive in seeking the attention of adults.

(9) Interferes with the work of others--e.g., imposes on other children and bosses them around; inappropriately insists on being admitted to a small peer group; destroys or takes away the sork or materials of other children.

(11) Does not follow directions--e.g., works or plays by self despite teacher's request that all children engage in group activity.

(13) Seeks to do things differently from others, even when own method is not effective--e.g., persists at tasks which are clearly impossible.

(16) Does not pursue task to completion--e.g., leaves task at hand for another or to wander about the room.

(18) Speaks out of turn; interrupts others; takes others' turn, etc.--e.g., refuses to wait his turn.

Administration of the Revised Classroom Behavior Checklist. As noted, two observers independently observed each child in a single classroom for five-minute periods, in sessions lasting up to two hours. The particular hours selected were scheduled well in advance and took into consideration the teacher's daily schedule of classroom activities.

These observers (different from last year's observers) were carefully trained during pilot phases of this year's activities. Their experiences yielded important qualitative information relevant to a replication study.

Each observer wrote a brief "narrative" as he was watching the particular child.[2] At the end of each day of observation all

---

[2]As noted in the Final Report for Interim Research, narratives allowed the opportunity for singular or unusual events to be recorded--events that were not anticipated, or even though they occurred but once, were thought to bear some significance to what we were after. This procedure was introduced in the light of the possibility that some unique aspect of behavior or some special circumstance could arise which was not or could not be anticipated by the behavior checklist, despite our careful pilot-testing in many classroom situations. Such narratives, we expected, would allow the final raters to take into consideration a qualifying, explanatory, or modifying circumstance which might alter the picture obtained only from frequency counts based on the checklist. In actuality, however, our observers have noted that only in rare instances did such narratives "clarify" the checklist information, and they were of little or no help when the final global ratings were done.

identifying data were removed from the narratives, and each observer independently reviewed his own narratives; the relevant behavior was then checked on the checklist (for each child) along a three-point scale--frequently, moderately often, rarely-never. A fourth category, no opportunity to observe, was also included.

It should be noted that the particular behavioral checklist scale employed during this year's replication differed slightly from that used last year, when a four-point scale was employed, treating rarely and never categories as separate scale points. This year, in keeping with the salient features of a checklist format, the observers as well as the investigators attempted to treat the emerging behavior more strictly along a nominal, yes or no, scale dimension. For example, the rarely-never category was checked if the described behavior did not occur during any given observation period. If the behavior occurred during the period, a check was made as to its occurrence (moderately often). If the behavior was particularly salient during the observation period by occurring repeatedly or spanning an extensive portion of the period, the frequently scale point was employed.

The formal data collection period began in October, 1969, and ran through most of December, 1969, in Public Schools 68 and 79. Subjects were independently observed by the two observers in their normal classroom settings. The total number of sessions and observations per child (across observers) ranged from 22 to 30. Observers made an equal number of observations for each child. No difficulties in either using the checklist or in the actual mechanics of classroom observations arose during the testing period.

The schedule for both observers was based on random selection of days of the week and random time-cluster sampling at each school

(modified only to balance the number of observations at each school). In addition, separate randomized lists of subjects were prepared for each observer for each day of observation. Departures from strict randomization of the ordering of subjects were necessitated, however, by the absence of any particular S. Whenever possible, a child who was not available at the scheduled time was observed later on a randomly determined basis.

B. Creativity Measures

After a thorough search of the literature, the creativity tests developed by Wallach and Kogan (1965) and Torrance (1966) were selected to be pilot-tested for modification or adaptation in some form or another depending on results of pilot-testing. This was decided after careful consideration of reliability aspects, correlation with other constructs, and similar questions. Accordingly, nine subtests from the above creativity batteries were administered in two sessions to seven third-grade children in P. S. 121 on East 103rd Street in New York City. (Not all subtests from the above authors' batteries were administered in this pilot-testing phase. For example, four of Torrance's verbal subtests were eliminated as they appeared to be too conceptual for our Ss.) Results of pilot-testing indicated that the productivity factor did not turn out to be a problem, as had been anticipated. With the exception of one of Torrance's tests, which was therefore eliminated, all the subtests yielded a wide range of scores.

Since both batteries were about equally reliable, and both (with the exception noted above) elicited a range of responses, the final selection of which battery to use was based on the following considerations: Torrance's figural tests were preferred to Wallach's and Kogan's as they require nonverbal responses, thus minimizing the verbal loading in

creativity. This scoring system also takes quality as well as quantity into account, whereas Wallach and Kogan's system does not.

Torrance's subtests were thus again pilot-tested with seven additional children in Public School 121 to work out additional details. Most important, the tests were carefully timed to determine whether Torrance's time limits were feasible with our Ss. It was decided to shorten the battery, since it was too lengthy for our purposes. The Construction subtest was thus eliminated since it was a one-item subtest which correlates .36 with the total battery as compared with correlations of .72 and .85 for the other two figural subtests with total battery.

Our final battery therefore consisted of four individually administered tests, two figural and two verbal, taken from the Torrance Tests of Creative Thinking (1966), Form A. The two figural tests are the Picture Completion test which requires S to draw pictures from 10 given incomplete sketches and the Lines task in which S is requested to draw different pictures on the basis of 18 pairs of parallel lines. The two verbal tests are the Product Improvement test requiring S to state the ways in which a stuffed toy elephant could be changed so that it would be more fun to play with and the Unusual Uses test in which S designates unusual uses for boxes. Pilot-testing with our population led us to modify slightly the four tests we used in terms of time limits, order of the tests, instructions, and, in one of the subtests, number of stimuli presented to S. Each of the tests yields either three or four scores: elaboration (amount of detail); fluency (number completed); originality (frequency of response); and flexibility (number of different categories in which the responses fall).

The testing of 39 subjects was completed in April, 1970. The ratio

of male (female) subjects to female and male testers was kept approximately equal. The testing sessions which included the Missouri Children's Picture Series along with the creativity battery, averaged about 40 minutes. The sequence of the tests was constant for all children and in the order of: Picture Completion, Product Improvement, Lines, and Unusual Uses. A figural test was chosen as the first task on the grounds that it would be more of an ice-breaker than a verbal one. Verbal tests were alternated with figural tests in the hope that the variety of different tasks following one another would reduce boredom. The MCPS followed the creativity battery. The MCPS was placed at the end of the session as it was considered a less fatiguing task. The instructions for the creativity tasks are presented below.

General Instructions. All timing of the tasks start at the end of the first instructions even in those cases where a second set of instruction is given. Instructions should be read in a conversational tone. The second set of instructions can be repeated as often as necessary. The stopwatch is to be used unobtrusively, though in sight of the S.

(1) Picture Completion

Materials: Pencil and test booklet.

The test booklet should be opened with both pages of the test in view. E says:

> WE ARE GOING TO PLAY SOME GAMES TODAY. HERE IS THE FIRST ONE. BY ADDING LINES TO THE INCOMPLETE FIGURES ON THIS AND THE NEXT PAGE, YOU CAN SKETCH SOME INTERESTING OBJECTS OR PICTURES. TRY TO THINK OF SOME PICTURE OR OBJECT THAT NO ONE ELSE WILL THINK OF. TRY TO MAKE IT TELL AS COMPLETE AND AS INTERESTING A STORY AS YOU CAN BY ADDING TO AND BUILDING UP YOUR FIRST IDEA. WHEN YOU'RE FINISHED YOU CAN TELL ME THE NAME OF EACH PICTURE.

(Start timing.)

If after 10 seconds the child has not started the task or looks confused, E adds the following instructions:

> TAKE YOUR PENCIL AND DRAW SOMETHING ON EACH OF THESE (E points to each of the first four stimuli) TO MAKE AN OBJECT OR PICTURE.

If the child starts with any other than the first stimulus, E says:

> DO THIS ONE FIRST, # 1, THEN THIS ONE, # 2, AND CONTINUE WITH # 3, AND THEN # 4. (E points to the stimuli.)

If S has finished the first page and does not realize that he is to continue on the second page, E tells him: DO THESE ALSO and places the second page in front of him. S is allowed 12 minutes to complete the task. E should note which of the 10 stimuli he is working on after 10 minutes have elapsed; or, if he is not working on any, the one he has completed. When S has finished making the pictures for the 10 stimuli or 12 minutes have elapsed, E says:

> TELL ME THE NAME OF EACH PICTURE. WHAT IS THIS ONE CALLED? (E points to the first.)

E then goes on to the other stimuli asking for the names if they are not given spontaneously and writes each one on the line provided.

(2) Product Improvement

Materials: Picture of the elephant.

E holds up the picture of the elephant and says:

> HERE IS A PICTURE OF A STUFFED TOY ELEPHANT OF THE KIND YOU CAN BUY IN MOST DIME STORES FOR ABOUT TWO DOLLARS. IT IS ABOUT SIX INCHES TALL (E indicates) AND WEIGHS ABOUT A HALF POUND. TELL ME THE CLEVEREST, MOST INTERESTING AND UNUSUAL WAYS YOU CAN THINK OF FOR CHANGING THIS TOY ELEPHANT SO THAT CHILDREN WILL HAVE MORE FUN PLAYING WITH IT. DO NOT WORRY ABOUT HOW MUCH THE CHANGE WOULD COST. THINK ONLY ABOUT WHAT WOULD MAKE IT MORE FUN TO PLAY WITH AS A TOY.

(Start timing.)

If because of the preceding task, S thinks he has to redraw the

elephant or if he looks confused or has not responded after 10 seconds, E says:

> TELL ME HOW YOU COULD CHANGE THIS STUFFED TOY ELEPHANT TO MAKE HIM MORE FUN TO PLAY WITH.

No probes are given. There is a 10 minute time limit. However, if S has been silent for one minute, E goes on to the next task without waiting for the 10 minutes to elapse.

(3) Lines

Materials: Pencil and crayons on a tray and test booklet.

When S has selected a pencil or crayon, the tray is removed. The test booklet is opened with both pages of the test in view. The test ends with stimulus 18 not stimulus 30.

E says:

> IN 14 MINUTES SEE HOW MANY OBJECTS OR PICTURES YOU CAN MAKE FROM THE PAIRS OF STRAIGHT LINES ON THESE TWO PAGES. THIS IS A PAIR, THIS IS A PAIR, AND THIS IS A PAIR. (E points to each of the first three pairs.) THE PAIRS OF STRAIGHT LINES SHOULD BE THE MAIN PART OF WHATEVER YOU MAKE. WITH PENCIL OR CRAYON ADD LINES TO THE PAIRS OF LINES TO COMPLETE YOUR PICTURE. YOU CAN PLACE MARKS BETWEEN THE LINES, ON THE LINES, AND OUTSIDE THE LINES--WHEREVER YOU WANT TO IN ORDER TO MAKE YOUR PICTURE. TRY TO THINK OF THINGS THAT NO ONE ELSE WILL THINK OF. MAKE AS MANY DIFFERENT PICTURES OR OBJECTS AS YOU CAN AND PUT AS MANY IDEAS AS YOU CAN IN EACH ONE. MAKE THEM TELL AS COMPLETE AND AS INTERESTING A STORY AS YOU CAN. AT THE END, YOU CAN TELL ME THE NAME OF EACH PICTURE.

E presents the tray and says:

> TAKE A CRAYON OR PENCIL. E then removes tray and starts timing.

If after 10 seconds, S does not start the task, E says:

> MAKE SOMETHING WITH YOUR PENCIL OR CRAYON STARTING WITH THESE TWO LINES: THIS PAIR (E points to # 1), THEN THIS PAIR (E points to # 2), THEN THIS PAIR (E points to # 3).

If S starts with any other than the first stimuli, E says:

DO THIS ONE FIRST, # 1, THEN THIS ONE, # 2, AND CONTINUE WITH # 3 AND # 4. (E points to the stimuli.)

S is then told to continue with the second page if he thinks there is only one page. The time limit is 14 minutes. E is to note the stimulus S is working on after 10 minutes and after 12 minutes. When S has finished or the 14 minutes are up, E asks S to tell him the name of each of the drawings and writes them on the lines provided.

(4) Unusual Uses

Materials. None.

E says:

MOST PEOPLE THROW THEIR EMPTY CARDBOARD BOXES AWAY, BUT THEY HAVE THOUSANDS OF INTERESTING AND UNUSUAL USES. TELL ME AS MANY OF THESE INTERESTING AND UNUSUAL USES AS YOU CAN THINK OF. DO NOT LIMIT YOURSELF TO ANY ONE SIZE OF BOX. YOU MAY USE AS MANY BOXES AS YOU LIKE. DO NOT LIMIT YOURSELF TO THE USES YOU HAVE SEEN OR HEARD ABOUT. THINK ABOUT AS MANY POSSIBLE NEW USES AS YOU CAN.

(Start timing.)

If the subject does not respond after 10 seconds, E says:

TELL ME AS MANY POSSIBLE USES FOR THE BOXES AS YOU CAN.

No probes are given. If there is a one minute silence, the task is ended. The time limit is 10 minutes.

C. The Sociometric Measures

The sociometric measures developed by our staff consisted of two parts, both given individually. In part (a) (ratings), each S is asked whether or not he likes each of the children in his class. In the choice test (b), S is asked to select two children for each of the following questions:

(1) Which two children would you like to sit next to?

(2) Which two children would you not want to sit next to?

(3) Which two children would you like to play with during recess?

(4) Which two children would you not want to play with during recess?

(5) Name the two children you like best, like very, very much.

(6) Name the two children you don't like at all--they really bug you.

Originally the choice test was a group-administered test of ten questions. S was to give the three choices for each question in order of his preference. The first rating scales piloted were also group-administered. They consisted of a page with four categories next to each child's name. The subject was to check the category which expressed the way he felt about the child.

In each of the four pilot testings at P.S. 80, East 120th Street, Manhattan, the tests were modified until the present format resulted. Such problems as misunderstanding of directions, reading and spelling problems, omissions of answers, and confusions in the answers were handled. Individual administration of the tests solved many of these problems.

Pilot-testing also disclosed the fact that there were frequent inconsistencies in response for some of the subjects. For example, an S might chose a particular child as liked on the ratings, but select him as a negative choice in the choice situation. To minimize inconsistencies, the tests were greatly simplified. Further, to make it possible for us to explore the reliability of the two sociometric measures, each part was given twice, in the following order: ratings; choice test; ratings; choice test.

The first administration of the tests was completed at the beginning of December, 1969. The second administration was completed in May, 1970.

(1) Instructions for Part (a)--Sociometric Ratings

Materials for test and retest: Cards with names of all classmates; answer sheets.

The test is to be administered in a corner of the classroom with the classmates in full view of the Ss.

Initial Instructions. THIS IS NOT A TEST. I AM GOING TO ASK YOU ABOUT CHILDREN IN YOUR CLASS WHOM YOU LIKE AND DON'T LIKE. NOBODY WILL SEE YOUR ANSWERS BUT ME. I AM NOT GOING TO SHOW THEM TO ANYONE. YOU ARE TO TELL ME ABOUT THE CHILDREN YOU DON'T LIKE AS WELL AS THOSE YOU DO LIKE.

E reads the names from the deck of shuffled cards and asks:

DO YOU LIKE HIM OR DON'T YOU LIKE HIM?

E records the child's rating (like or don't like) of each classmate on an individual answer sheet which contains the names of all the children in his class.

Retest Instructions. WE'RE GOING TO DO THIS AGAIN. (S may see the scoring sheet on the initial test, but not on the retest.)

(2) Instructions for Part (b)--Sociometric Choice

Materials for test and retest: Answer sheets.

Initial Instructions. I AM GOING TO ASK YOU SOME QUESTIONS ABOUT CHILDREN IN YOUR CLASS WITH WHOM YOU WOULD LIKE TO DO THINGS AND NOT LIKE TO DO THINGS. NOBODY WILL SEE YOUR ANSWERS. YOU MAY CHOOSE SOMEONE EVEN IF HE IS ABSENT TODAY. YOU MAY CHOOSE A CHILD MORE THAN ONCE IF THAT IS THE CHILD YOU REALLY WANT. LET US SAY I HAD A GOOD GRIEND NAMED MARILYN (OR JEROME) IN THIS CLASS. I WOULD PROBABLY WANT TO SIT NEXT TO MARILYN (JEROME) AND PLAY WITH HER (HIM) AND WALK TO SCHOOL WITH HER (HIM). LET'S SAY I DIDN'T LIKE A GIRL (BOY) CALLED TOBY (MARSHALL) IN THE CLASS. I WOULD NOT WANT TO SIT NEXT TO TOBY (MARSHALL), NOT WANT TO PLAY WITH HER (HIM) AND NOT WALK TO SCHOOL WITH HER (HIM). (Male examiners use male example.)

E reads the following questions:

(1) WHICH TWO CHILDREN WOULD YOU LIKE TO SIT NEXT TO?

(2) WHICH TWO CHILDREN WOULD YOU NOT WANT TO SIT NEXT TO?

(3) WHICH TWO CHILDREN WOULD YOU LIKE TO PLAY WITH DURING LUNCH, RECESS, OR FREE TIME?

(4) WHICH TWO CHILDREN WOULD YOU NOT WANT TO PLAY WITH DURING LUNCH OR FREE TIME?

(5) NAME THE TWO CHILDREN YOU LIKE BEST, LIKE VERY MUCH.

(6) NAME THE TWO CHILDREN YOU DON'T LIKE AT ALL; THEY REALLY BUG YOU.

Retest Instructions. E says: NOW I AM GOING TO ASK YOU THE SAME QUESTIONS AS BEFORE. IT'S ALL RIGHT TO CHOOSE THE SAME CHILDREN AS LAST TIME IF THAT'S WHOM YOU WANT.

E repeats the above questions.

D. The Missouri Children's Picture Series (MCPS)

At the end of Chapter 2, we discussed some of the rationale behind using this instrument to assess certain personality variables. As indicated above, the MCPS was administered immediately after the creativity battery. This order was established in pilot testing in an attempt to maintain a high motivational state during the testing session. Pilot results were based on explorations with third-grade children in Public School 121 during the winter of 1970. Pilot testing also disclosed that the MCPS is an objective, enjoyable, brief instrument which is easily administered on an individual basis to a sample of elementary school children, requiring, on the average, about twenty minutes of testing time.

Formal testing of the Ss was completed during the month of April, 1970. Because the instrument was administered in the same session as the creativity battery, the ratio of male (female) Ss to female and male testers was kept approximately equal.

Instructions.

Materials: Set of 238 standard MCPS cards; scoring sheets.

The set of 238 cards is shuffled after each use. The cards do not have to be in numerical order. The set of cards is placed by the E in front of the child, picture sides up. The colored card on the bottom of the set is removed and placed to the right (the S's right) of the deck of cards. S is then asked to pick up the card which is on top of the deck and he is told:

> ALL THESE CARDS HAVE PICTURES ON THEM. I WANT YOU TO LOOK AT EACH PICTURE AND SEE IF IT LOOKS LIKE FUN TO YOU. IF A PICTURE LOOKS LIKE FUN TO YOU, PUT IT ON THE (YELLOW/BLUE) CARD, HERE ON THE RIGHT. IF IT DOES NOT LOOK LIKE FUN TO YOU, PUT IT ON THE OTHER SIDE OF THE CARDS. NOW LOOK AT THIS FIRST PICTURE. DOES IT LOOK LIKE FUN TO YOU?

If S says that it does look like fun, he is told:

> OKAY, PUT IT HERE, ON THE RIGHT, ON THE (YELLOW/BLUE) CARD. THAT IS WHERE YOU WILL PUT ONES THAT LOOK LIKE FUN TO YOU. THE ONES THAT DO NOT LOOK LIKE FUN TO YOU, YOU WILL PUT HERE, ON THE OTHER SIDE OF THE PILE OF CARDS, ON THE LEFT (E points).

If S says that the first picture does not look like fun, E says:

> OKAY, PUT IT HERE, ON THE LEFT, ON THIS SIDE OF THE PILE OF CARDS (E points). THAT IS WHERE YOU WILL PUT THE ONES THAT DO NOT LOOK LIKE FUN. THE ONES THAT DO LOOK LIKE FUN TO YOU, YOU WILL PUT HERE, ON THE (YELLOW/BLUE) CARD, ON THE OTHER SIDE OF THE PILE OF CARDS, ON THE RIGHT.

S is now told:

> PICK UP THE NEXT CARD AND LOOK AT IT. DOES IT LOOK LIKE FUN TO YOU OR DOESN'T IT LOOK LIKE FUN TO YOU?

If S says that it looks like fun, he is told:

> OKAY, PUT IT ON THE RIGHT THERE, ON THE (YELLOW/BLUE) CARD. (After this has been done): WHERE WOULD YOU HAVE PUT IT IF IT HAD NOT LOOKED LIKE FUN?

If S says that the second picture does not look like fun, he is told:

> OKAY, PUT IT ON THE OTHER SIDE OF THE PILE OF CARDS,

ON THE LEFT. (After the card has been placed correctly): WHERE WOULD YOU HAVE PUT IT IF IT DID LOOK LIKE FUN?

If the E feels that the S is confused about the instructions or the task at hand, he should then continue the instructions with a third card from the deck. Generally, if at anytime during the sorting the examiner feels that the S is confused or faltering, he should repeat these instructions.

The E continues in this way until he is sure that S understands what is wanted of him, and he observes S at least long enough to be sure that he is able to attend to the job. In the event that the S is very distractible, E may have to repeat the question for each card in the set.

Sometimes the S will ask questions when he is unsure of himself, or when he is suspicious of the testing situation, or when compulsivity or concreteness interferes with easy decision-making. These questions should be responded to either with the original instructions or else with amplifications which keep to the spirit of the instructions and do not influence the S to respond one way or the other. Following are some examples of questions and possible replies:

"What if I can't decide whether it looks like fun or not?"

JUST LOOK AT IT AND DECIDE ONE WAY OR THE OTHER. SEE IF IT LOOKS JUST A LITTLE BIT MORE LIKE FUN OR IF IT LOOKS JUST A LITTLE BIT MORE LIKE IT IS NOT FUN. THEN PUT IT ON THE PILE AND FORGET ABOUT IT.

"Sometimes it's fun and sometimes it isn't. What should I do?"

JUST LOOK AT IT AND DECIDE IF IT IS MORE LIKE FUN THAN NOT LIKE FUN, OR THE OTHER WAY AROUND. DECIDE ONE WAY OR THE OTHER AS BEST YOU CAN AND THEN FORGET ABOUT IT.

"This used to be fun when I was smaller. Do you want me to say if it's fun now or if it used to be fun?"

I WANT YOU TO SHOW IF IT LOOKS LIKE FUN TO YOU OR IF IT DOES NOT LOOK LIKE FUN TO YOU. DECIDE ONE WAY OR THE OTHER AND THEN GO ON TO THE NEXT ONE.

(From a girl): "This is fun for boys but not for girls. Should I show if it's just fun, or if it's fun for girls?"

I WANT YOU TO DECIDE WHETHER IT LOOKS LIKE FUN TO YOU OR WHETHER IT DOES NOT LOOK LIKE FUN TO YOU.

When the above is completed, E immediately records the sort on the scoring sheet in accordance with instructions described next in the chapter.

E. Teacher's Ratings

Teachers' individual ratings for each child in the sample were obtained in the Fall of 1969, and again in the Spring of 1970. Teachers were requested to rate, globally, on the behavior dimensions of independence, dependence, and pseudoindependence. For this purpose, the global rating scales employed for the Global Behavior Checklist ratings were also employed for the Teacher's Ratings (see end of next section).

In the Fall of 1969, meetings were held with the teachers in an attempt to elicit their cooperation for the year's study, as well as to extensively discuss with them the definitions of those behaviors characterized as independent, dependent, and pseudoindependent. These meetings were used specifically for imposing some form of face validity upon the rating scales employed. By ensuring a common definition of the specific behaviors tapped, at least one prerequisite for valid rating scales was fulfilled.

Two one-hour meetings were held with each third-grade teacher and his (her) assistant teacher, as well as the grade supervisor for that particular school. Teachers were provided with our working definitions

of behavioral categories (see Section A, this chapter); however, no information was given as to the use of these definitions in the context of the Revised Behavior Checklist or Global Checklist Ratings. These discussions were of a seminar-type, designed to inform the teachers of the general purposes of the study, and to elicit their optimal cooperation and understanding for the year's work.

Provisions for considering the intensity of the observed behavior (high or moderate) as well as the teacher's confidence in making the rating were included in the rating scales. Teachers were allowed three days (both in the Fall and Spring) to make their ratings. Although in the Fall teachers were told of the general schedule for the year's work, they were not told specifically of the retest period for their ratings which would follow in the Spring.

F. Global Checklist Ratings

As in last year's study, two independent raters (the Principal and Co-Principal Investigators) globally rated the Ss on the basis of two summary sheets of the Revised Classroom Behavior Checklist observations (one for each observer) which contained total frequencies for each scale point for each item on the behavior checklist. As noted, it was not possible to use the "narratives" (comments by the observers on the child's classroom behavior with regard to anything salient that might have occurred) in this process, since they did not contain sufficiently complete or consistent information from child to child to warrent their use.

Rating Scale for Teachers and for Global Checklist Raters

Rater_____ IDS Code No._____

Check one only:

I--Independent Highly_____ Moderately_____

P--Pseudoindependent Highly_____ Moderately_____

D--Dependent Highly_____ Moderately_____

X--Can't rate_____

Degree of confidence in rating--check one:

3_____high 2_____medium 1_____low

The next chapter describes the scoring methods we used, reliability findings, and the results of data analyses in connection with the hypotheses and expectations of our research.

Chapter 4

Scoring Methods, Data Analyses, and Findings

This chapter is divided into two sections. The first presents a detailed description of the various scoring techniques we used for purposes of data analyses and data reduction. This section also presents reliability information concerning our scores and methods. The second section presents findings in connection with the major purposes of the investigation: the overall hypotheses and expectations as presented in Chapter 1.

I--Scores and Reliability

A. Revised Classroom Behavior Checklist

Checklist frequency scores. The Checklist comprises items which fell into three categories of Independence, Pseudoindependence, or Dependence. Observations of these behaviors for each day of observation for each child (over a period of time) were tallied for the different frequency points: frequently, moderately often, and rarely/never. It should be mentioned again that these particular scalepoints differ slightly from those used in last year's work.

In order to derive individual scores (ultimately a Z-score based on each dimension's distribution), for each subject a frequency count was made of the number of tallies per item for each of the frequency points for each of the observation sessions, with data for each observer treated separately. The can't rate point was excluded in this procedure. Totals of these counts for each child per item across all of the observation sessions were then calculated. These computations yielded the total number of "ratable" time-sampling observations made for each child

for all sessions for each observer.

To obtain a score for each child, weights were assigned: frequently--3; moderately often--2; and rarely/never--1. (Higher scores for any item reflect "more" of the particular behavior observed, i.e., greater independence, greater pseudoindependence, greater dependence.) These weights were then multiplied by the number of observations for the scale point, yielding a total score for that scale point. These scores were then totaled across scale points. Totals were then divided by the number of ratable observations to yield a mean score for the child, thereby eliminating irrelevant variance resulting from unequal numbers of observations. Unlike last year's format, the addition of a constant to this mean score was not required to eliminate zero scores because of the weighting procedure employed. Last year's format of four scale points employed weights ranging from three to zero. This year's procedure eliminated the zero weight to avoid zero scores for any subject.

Inter-observer reliability. Following a procedure developed during last year's work with the Checklist, the foregoing mean scores were paired for each independent observer and Pearson correlation coefficients were then computed for these paired scores. The coefficients thus obtained indicate the amount of agreement between observers. These coefficients are reported in Table 3. Note that the n per item varies across items because the scores employed for reliability purposes are based only on observable (ratable) time samplings, a variable factor. Note also that Item 1 in the Dependence group (withdraws in the face of difficulty) was not sufficiently represented in observations to warrant computing a correlation coefficient. The same is true for Item 13 in the Pseudoindependence group (seeks to do things differently from others, even when

own method is not effective).

In order to establish the basic score for each S (Z-scores), items falling below an arbitrarily determined coefficient level of .15 were eliminated from further consideration. These are Items 15 and 19 from the Independence group (r = .10 and .04 respectively); and Items 3 and 6 from the Pseudoindependence category (r = .13 in both instances). Of the remaining items, only two (Items 10 and 16) in the final pool of items fall below the .25 level.

Z-score derivations. The final pool of items for Z-score derivations thus includes 13 items. These are the underlined items in Table 3.

As developed in last year's work, the "scoring system" for the Checklist was based on computing Z-scores (separately for each school) for each S. Unlike last year, however, these scores were not based on the distribution of scores for each item. Rather, the distribution of scores (which yielded a Z-score for each S) consisted of mean scores for each behavior category for all items within that category. That is, a mean category score was derived for each behavioral dimension (independence, pseudoindependence, dependence) which was based on each S's score for each item and each observer.

Within each school, Z-scores were calculated for each category for each S. Table 4 presents the Ss ranks in all three categories for both schools combined. Means of Z-scores for each school for each category are also presented in Table 4. Were there only chance differences between schools in the distribution of the S's scores, then mean of ranks for each school for each category would be 20 or as close to 20 as possible. In actual fact, as Table 4 indicates, there is little difference in mean rank between schools within each designation.

A "rule-of-thumb" was then devised for designating the Ss for purposes of data analyses on the basis of their ranks on the three Z-score distributions. For the current sample, the optimal cut-off points for establishing the Ss' designations appeared to be the following Ss with ranks of 20.5 through 40 (at the median or above) for the independence category are to be regarded as "independents" provided that their Z-scores for the other categories fall below the 55th percentile for these respective distributions, that is, fall in ranks 1 through 22 on both dependence and pseudoindependence distributions. Subjects with ranks in the independence category of 1 through 18.5 (below the median) are to be regarded as "nonindependents" provided that they fall in the top 45th percentile (ranks 23-40) for one or the other, or both of the two remaining category distributions. On the basis of these criteria the designations of the subjects resulted in classifying 10 subjects as "independent," 14 as "nonindependent," and 16 as "mixed" types. Of these 16 "mixed" cases, there are 11 subjects who rank above the median on independence as well as above the median on at least one of the other behavioral category distributions. Subjects are listed by their Z-score rank for each category distribution as well as their overall designation in Table 5.

B. Creativity Measures

Scores and reliability. The four creativity subtests used in the current study yielded a total of seven subscale scores. The figural tests (Picture Completion and Parallel Lines) were scored for fluency--the number of unrepeated responses emitted; elaboration--the amount of detail present in each response; originality--the uniqueness of each response; and flexibility--the total number of subject categories into

which responses fell. The verbal subtests were scored only for fluency, originality, and flexibility as suggested by established scoring procedures (Torrance, 1966). The complete scoring procedures we employed for the creativity measures are described in detail in Appendix B.

The Ss' responses were scored by two independent scorers, according to the scoring instructions provided by Torrance (op. cit.). It should be noted, however, that in some instances, Torrance's scoring system required some modification because of ambiguous directions in the scoring manual itself. For example, a fluency score on the Parallel Lines subtest is typically derived from the total number of non-repetitive drawings provided by the subject. In the present instance, the scorers found it necessary to indicate more clearly than in standard procedures the differentiation between repetitive and non-repetitive responses. A second type of modification was necessary because of specific sample characteristics. Torrance's scoring for originality, for example, was based upon the distribution of responses in a predominantly white, middle-class sample. In the present case, scoring for originality included not only Torrance's standardized categories, but also those derived from the current, ghetto sample.

For each subscale in the creativity battery, total scores were derived on the basis of tallies representing the number of times that particular response characteristic occurred for each item during the test administration. As noted, two independent scorers computed total scores for each child and each subscale. To index the reliability of the present scoring procedures, inter-scorer reliability coefficients were computed. For this purpose, simple Pearson correlational methods were employed. As reported in Table 6, a high degree of agreement was

found for the scoring procedures developed.

To develop scores for purposes of data analyses, the scores derived by the independent scorers were reviewed, and discrepant score assignments were discussed in order to reach agreement on each of the seven subscores for each child.

For each section of the battery, that is, Figural or Verbal, a final fluency score was derived which indexes the fluidity of unrepeated responses for each of the subscales. For present purposes, this score was used as a basal measure in deriving all other scores. This measure provided a basis for determining the extent to which flexible, original, or elaborative responses actually accounted for a response which could be attributable to the individual's creativity. That is, the basal measure partialed out variance attributable mainly to fluidity in responding. Therefore, although fluency scores represented a tally of the number of emissions for each subscale, the remaining scores, those of flexibility, originality, and elaboration (for the Figural tests) represented ratio scores of number of salient emissions corrected for (divided by) the total number of emissions. All ratio scores, for analysis purposes, were multiplied by a constant (+10) for ease of data reduction and comparability to the fluency scores derived.

C. Sociometric Measures

Sociometric ratings: the Like-Dislike scale. The number of times each child was chosen as liked or disliked by every other child in his class was tallied for each administration (test-retest) in each test administration period (Fall, Spring). To index consistency of choice and develop a reliable scoring procedure for this sociometric measure, a contingency table was devised reflecting the consistency of each S's

choice on each administration (test and retest) of the like-dislike scale. That is, for each S and each item of the scale the various combinations of selecting like or dislike on the two administrations of the scale were tallied. Taken across children, the combinations resulted in a four-fold table: each axis represented a respective time (test or retest) of administration with the alternatives of like and dislike.

To index the degree of association between these two axes a chi-square analysis was employed, yielding a phi (Ø) coefficient. The N for each analysis represents the total number of choices made per school and for the total sample. That is, N represents the total number of times each child rated every other child, summed across children.

This procedure, then, formed the basis for developing a reliable scoring format. As indicated in Table 7, the association between choices on the first and second administrations of the sociometric test is high, significant, and positive for each separate testing period. There seem to be barely any school differences in these findings.

Rating scale: scoring procedures. The scores used for hypotheses-testing involved procedures developed for the reliability analyses by employing scores based only on those instances where there was complete consistency in choice when test and retest situations were compared. This was repeated for each of the test periods (Fall and Spring).

A notation was made if the child chosen was rated as liked or disliked on both test and retest administrations of the rating scale. The consistent choices thus derived for each child were summed (tallied) for all the children, resulting in a single score representing the number of times that child was consistently chosen as liked or disliked by

individual members of the class. A child's final score, in this case, is taken as the ratio of the number of times he is chosen as liked divided by the total number of times he was consistently rated (as liked or disliked) by the other children. This popularity index, then, controlled for the unequal number of cases where inconsistencies in choices appeared among the children.

A check on sex differences in rating behavior was made to control for score variance unattributable to those behavioral correlates which we are studying. To investigate rating behavior, or popularity, as a function of the sex of the child chosen and the sex of the rater, *t*-tests were computed for each sex group as a function of the sex of the rater. That is, for boys, a *t*-test was run comparing the mean ratio scores for girls' choices and boys' choices. The same was done for the girls in the sample. These tests were computed for each school separately. The analyses yielded highly significant results for each testing (Fall and Spring) in that there was a very strong association between sex of rater and sex of child chosen. Table 8 presents these findings.

To control for this score variance, a child's score was taken as that ratio based only on his appropriate sex group. If the subject was male, for example, his score was the ratio of the number of times he was chosen as liked by the boys in the sample, to the total number of ratings made of him by the boys in the sample. These scores, then, were the scores employed for analyses purposes.

Rating scale change scores. Inquiry was made into the nature of change in sociometric status for the two separate testing periods (Fall and Spring) for the rating (like-dislike) scale. For this purpose, a simple rank order method was employed in designating S's status on each

administration of the test. This method was employed specifically to avoid change scores based on unequal numbers of cases per child per administration. Further, it was felt that quantitative change scores would not take into account initial (Fall) score level. Finally, any apparent ceiling effect for choice scores would confound an index of change. It was decided, therefore, to study (by pattern analysis) only those extreme cases of change and their source. For this purpose, equivalent cut-off points were employed for each test period's distribution of scores and Ss were designated as high or low in popularity. These points were specifically the upper and lower 30th percentiles of each distribution for each test administration period.[3]

Sociometric choice. In the choice situation, the Ss were asked six questions involving the selection of two children they would like best or least in various types of situations (see pp. 53 and 54 of this report). To explore the reliability of the Ss' responses, agreement with regard to the test-retest administrations was weighted according to degree of consistency in the following way: total inconsistency = 0; partial consistency = 1; and total consistency = 2. For each of the six items in this part of the sociometric situation, the number of agreements, partial agreements, and disagreements (from test to retest) was tallied for all Ss in the sample. Because our interest was in determining the degree of consistency of choice from test to retest, the weighted category (2)--complete consistency--was tested (chi-square method) against weights (0) and (1)--total inconsistency and partial consistency. By chance alone, one-third of the cases would fall into weighted category (2), it should be noted.

Table 9 presents the results of chi-square tests of association

[3]Since only 3 cases of marked increase and 3 cases of marked decrease in popularity were found, further examination of these instances of sociometric change was not attempted.

for test-retest administrations of the sociometric choice situation for both Fall and Spring testings. These were computed for both schools combined to maximize the sample size on which these analyses were based. For the analyses of data based on the Fall test administration, Table 9 indicated that the distributions for items (1), (4), and (5) differed significantly from a chance distribution. Interestingly, for items (1) and (5)--and these are positive choice situations--the significant finding is based on agreement of choice from test to retest. Item (4), however--a negative choice situation--yielded an unexpected significant finding: that is, complete agreements from test to retest are fewer than expected on a chance basis alone. Because of these findings, only items (1) and (5) of the choice situation were scored: for each of the two administrations of the test (during the single session), the raw score consisted of a tally of the number of times the S was chosen in response to items (1) and (5).

Table 9 indicates that in the Spring, only item (1) demonstrated significant consistency in test to retest trials. However, for scoring purposes, items (1) and (5) were both used to index the S's responses in the sociometric choice situation. Item (5) was included in order to provide a wider range of situational descriptions in indexing choice behavior. It is interesting to note that the questions in this situation which involved positive choices (e.g., Which two children would you like to sit next to?) were more stable from test to retest than those items which involved negative choices (e.g., Which two children would you not want to sit next to?). Again, a tally was made of the number of times in the test and retest situation each child was chosen by other children in items (1) and (5). These represented the S's raw scores in this situation.

Sociometric choice scores based on items (1) and (5) were tallied separately for each sex. This distribution indicated that in the Fall the Ss chose a member of the opposite sex only 15% of the time for these items. In the Spring the percentages were slightly lower. For item (1), Ss chose a member of the opposite sex 10% of the time, and for item (5), 12% of the time. Sociometric choice scores, then, were based on the total number of times a child was chosen by Ss of the same sex (on both test and retest measures) for each test period, Fall and Spring. These scores were then coded according to the frequency distributions of the number of choices which appeared for each item as presented in Table 10. Based on the combined sex distribution, a nominal scale score was designated to reflect popularity status. Those subjects who received no choices were designated as lower in popularity on these items. Those subjects who received two or more choices were considered higher in popularity status on each item. For the total sample on item (1), for the Fall, the number of lows was 16 and the number of highs was 10. In the Spring, 20 Ss were designated low in popularity and 15 were designated as high. For item (5), in the Fall, the number of lows was 19 and the number of highs was 10. In the Spring, 22 Ss were designated as low, and 13 as high.

To investigate the relationship among sociometric measures, phi coefficients were computed on the basis of the dichotomous scoring system developed above. The two sociometric choice items employed relate positively and significantly on this basis. In the Fall, the phi coefficient between items (1) and (5) of the sociometric choice test was .91 (chi square = 19.92, $p < .001$); in the Spring, this coefficient was found to be .73 (chi square = 17.06; $p < .001$). It should be noted

further that scores for each of the items were stable across test periods. The stability coefficient between Fall and Spring test periods expressed as phi for item (1) was .67 (chi square = 10.37, $p < .01$); and that for item (5) was .58 (chi square = 8.60; $p < .01$).

The relationship between each sociometric choice item and the sociometric rating score was also investigated, employing the point biserial correlation coefficient for each sociometric choice item and the sociometric rating in the Fall and Spring. Again, these correlations are significantly positive. In the Fall, the point biserial correlation coefficient between item (1) and the sociometric rating was found to be .54 ($t = 3.258$, $p < .01$, two-tailed). The correlation between item (5) and this rating was found to be .55 ($t = 3.589$, $p < .01$, two-tailed). In the Spring, the correlation of item (1) with the rating was found to be .36 ($t = 2.138$, $p < .05$, two-tailed). And the correlation between item (5) with the rating was found to be .39 ($t = 2.688$, $p < .05$, two-tailed).

D. Missouri Children's Picture Series

Scores and reliability. Each picture has a number in the lower right-hand corner of the card. They are numbered from 1 to 238. The responses are recorded on a scoring sheet which contains 238 numbered spaces. Only the does-not-look-like-fun responses are recorded. Transparent templates, provided for each subscale, are used for deriving raw scores on each subscale. The templates contain X and O markings defining key items for each scale. An X on a template means that on that scale the item is scored if it was placed in the not-fun pile (and therefore has an X marked in the appropriate space on the recording sheet). An O on a template means that on that scale the item is scored if it was placed by the child in the fun pile (so that the corresponding space on the recording sheet is blank). Each X or O response has a weight of 1, and the raw score for a particular scale is the number of spaces on the recording sheet which match those on the scoring template.

The raw scores were converted into T-scores with the use of conversion tables provided by the test developers. These tables were derived from the MCPS results on the 3877 school children in the normative sample. There are separate tables for boys and for girls at yearly intervals from ages 5 through 16. The following subscales of the instrument have been scored and are employed in current data analyses: Maturity, Conformity, Inhibition, Aggressivity, and Hyperactivity.

Research with MCPS has been extensive and reliability as well as validity data have been reported. It should be noted (Sines, Pauker, & Sines, 1968) that the items which have been employed in all of the subscales of the MCPS show significant and positive discrimination among known test samples. Furthermore, these subscales have been investigated

in terms of internal consistency of items, ten-day retest reliability, and six-month retest reliability. These data are reported for each age-group and sex separately.

Because sex of subject enters into the scoring procedures, for current data analysis purposes, subscale scores were held separately for each sex. That is, an S's status on any particular subscale depends only on the distribution of such subscale scores for that sex group in our current sample. A median cut-off was established for each subscale distribution for each sex, and S's status--high or low--was established. Cases were eliminated where the S's rank fell at the median of the distribution of scores. The total possible N for this instrument was 39, 18 females and 21 males. For each subscale, the frequencies thus obtained are presented in Table 11.

E. Teachers' Ratings

To assess the stability of teachers' ratings, a contingency coefficient was computed relating Fall and Spring designations for each S (see Table 12). This coefficient (C = .66) is highly significant as determined by the chi square method (chi square = 27.31, $p < .01$, 4dfs). This analysis was based on combined school data in order to permit a more powerful statistical test to be performed.

The N for the foregoing analysis was 35, rather than the larger total sample since cases were eliminated where "Can't Rate" ratings were employed either in the Fall or Spring, or both. As shown in Table 13, five such cases were found. (For one of these cases, the "Can't Rate" category was employed both in the Fall and Spring.)

As can be observed in Table 12, teachers shifted more frequently in Public School 79 than in Public School 68. In the latter school,

only one change was found: a rating was changed from Independence in the Fall to Dependence in the Spring. Although the total number of cases rated Independent in the Spring is reduced by one, there is no equivalent increase reported for Dependence. The reason for this is that one case rated Dependent in the Fall was given a "Can't Rate" rating in the Spring, thereby reducing this category by one in the Spring tally.

For Public School 79, ratings for six cases were changed from Fall to Spring periods. Of the Fall Dependence cases, one rating was changed to Independence in the Spring. For the Fall Pseudoindependence cases, two of these were changed to Independence ratings in the Spring, and three were changed to Dependence ratings in the Spring. No changes were reported for cases rated as Independent in the Fall for Public School 79. For Public School 79, ratings for six cases were changed from Fall to Spring periods.

F. "Global" Checklist Ratings

After the global ratings based on the checklist were completed, interrater reliability was determined, as in last year's study, on the basis of Cohen's weighted kappa (1968), a coefficient for nominal scales. This measure not only corrects for chance agreement, but also allows for differential weighting of disagreements according to the degree of gravity of such disagreements. Therefore, it was again possible to weight scale-point disagreements (i.e., highly, moderately) less heavily than categorical disagreements (i.e., independent, pseudoindependent, dependent). Those weights derived for reliability purposes in last year's sample were once again employed in the current study. Table 12 in Appendix A contains the matrix of the weights employed for this

procedure. These weights range from 0 to 4. The larger the weight, the greater the disagreement; 0 indicates no disagreement, and 4 indicates maximum disagreement.

Table 14 (current report) shows the distribution of the global ratings made by the two independent raters in the current study. These data yielded a reliability coefficient (weighted kappa) of .38 based on 34 cases, taking into account scale points as well as category. This coefficient resulted in a $z = 2.92$ ($p < .002$, two-tailed). This coefficient of agreement, corrected for chance agreement between the raters, differs significantly from zero. It can therefore be concluded that the raters were able to demonstrate agreement in designating Ss in the current sample into three relevant classifications on the basis of the summary checklist data.

II--Major Findings

This section presents, as noted, findings in connection with the major purposes of the investigation: the overall hypotheses and expectations as presented in Chapter 1 (see pp.5-8).

A. Revised Classroom Behavior Checklist

Table 15 indicates that Z-score Checklist designations and Global Checklist ratings are highly related ($p < .01$). Below are findings more specifically related to interrelationships between independence (as measured in our study) and other variables in which we are interested.

Relationship to Teachers' Ratings. One major hypothesis explored concerns the relationship of independently designated classifications of the Ss in regard to the basic dimensions of this study (independence, pseudoindependence, and dependence) as derived from classroom observations in the Fall and teachers' ratings made at the beginning and at the end of the school year as to these same dimensions. Our expectation was that the teachers' ratings (made in the Spring) would show a higher positive correlation with the Fall Checklist designations than would the Fall teachers' ratings with the Fall Checklist designations. Our assumption was that as teachers got to know their children better, their judgments would become more reliable and valid.

Table 15 presents results for the foregoing analyses. It can be seen from these findings that Z-score designations tend to be positively associated with teachers' ratings in both the Fall and the Spring, and that, in accordance with our expectations, the positive relationship is a more confident one for the Spring teachers' ratings. Neither Fall nor Spring Teachers' ratings were significantly associated with global Checklist ratings, it should be noted. (These findings were based on

chi square analyses using four-fold tables with one degree of freedom.)

Relationship to Sociometric Ratings and Choices. Designations (Z-score and global ratings) of independence were compared to Fall and Spring sociometric nominations (ratings of like and dislike as well as of choice in various types of situations). Our expectations were that the dependent child will be less likely to be chosen and the pseudo-independent child least likely to be chosen in various sociometric contexts at the beginning of the school year but that no clearcut relationships will be found with scores for independence. On the other hand, in the Spring, sociometric scores will be even less likely to be positively related to designations of dependence; also, independent children will most often be chosen and/or rated as liked. It was also expected that pseudoindependent children will be nominated as least chosen or liked when Spring relationships were explored.

Tables 16 and 17 present findings with regard to these designations and sociometric ratings and choices for both the Fall and the Spring. As can be seen in Table 16, Z-score classifications and sociometric status as determined by the rating method were not significantly related in either Fall or Spring testing periods. Z-score designations and sociometric choices (items 1 and 5) also were not significantly associated with each other either in the Fall or the Spring (see Table 17). However, as Table 16 indicates, global ratings tended to be associated with Spring sociometric ratings (like-dislike). Global ratings bore no relationship to sociometric choices either in the Fall or the Spring, it might be noted. The one trend described above, regarding like-dislike ratings, is in the expected direction: in the Spring, children rated as more popular tended to be among those globally rated as independent by

the two independent raters on the basis of the checklist observations. In terms of the hypotheses noted above, then, it can be said that our expectations were in part confirmed concerning the popular sociometric status of independent children in the Spring, rather than in the Fall. Note, expectations concerning the lack of popularity of the pseudo-independent children could not be explored in the light of the nature of data distributions. That is, we found it possible to work only with independent vs. non-independent Ss, in the light of the small numbers of pseudoindependent Ss designated.

Relationship to Creativity Scores. To investigate the relationship between sex of subject and various creativity scores, t-tests were run for each subscale of the creativity battery as determined by sex. The results of these analyses are reported in Table 18. As can be observed, there is virtually no difference between males and females on these subscales with the exception of the Figural Originality subscale ($t = 2.13$, $p < .05$). On this subscale, males score significantly higher tha females. For subsequent analyses, therefore, males and females were treated separately for this particular scale. Further, as can be observed in Table 18, there are virtually no school differences for any of the scores on the creativity battery subscales. Therefore, schools were combined in order to provide greater statistical power.

Checklist scores were then compared to various measures of creativity employing t-tests. Tables 19 and 20 present the p values for t-test analyses of significance of differences between means for various creativity scores and dichotomous groups of independent and non-independent Ss as defined by the Z-score designations and global ratings respectively. We had expected, it should be noted, that creativity scores

would be positively related to various designations of independence but negatively related to designations of non-independence (pseudoindependence and dependence) derived from the classroom behavior checklist.

An examination of Table 19 shows one significant finding and a strong trend. The significant finding is in the unexpected direction: non-independent Ss as defined by Z-score classifications had significantly higher means on the Verbal Flexibility score than independent Ss. There was a strong trend in the expected direction, however, for independent Ss designated on the basis of Z-scores to have higher Verbal Fluency scores than non-independent subjects. Significant relationships were not found for the other creativity scores and Z-score designations.

Table 20 indicates one significant finding, again, not in the expected direction: non-independent Ss as determined by global ratings had significantly higher means for Figural Flexibility than independent Ss. Other findings with regard to global ratings and creativity scores were non-significant.

Overall, though there is a trend for higher means on various creativity scores to be obtained by the independent Ss (designated by Z-score as well as global ratings), the two significant findings obtained were not in the expected direction. It can be concluded that our expectations concerning the positive relationship of independence and various creativity scores in this group of children were not met.

Relationship to Missouri Children's Picture Series Subscales. With regard to the personality variables presumably measured by the MCPS, we had expected that higher scores on subscales prejudged to be representative of "healthy" or "good" self-concepts would be obtained for significantly more children designated as independent than as pseudoindependent

or independent on the basis of the classroom Checklist and that the reverse would also be true (scales representative of "poor" self-concept would be obtained for significantly more of the children designated as non-independent). Expectations were not held, however, for scales for which this type of a priori judging was not possible.

To investigate the foregoing relationships, ten 2 x 2 chi square analyses were performed (Yates corrected)--for each of the five MCPS subscales dichotomized as high vs. low scores and independence vs. non-independence categories for both global and Z-score checklist designations. The subscales are: Conformity, Maturity, Aggressivity, Inhibition, and Hyperactivity. No significant relationships appeared among any of these variables, as can be seen in Table 21. Our expectations regarding independence ratings (Checklist) and MCPS scores were therefore not met, it can be concluded.

B. Teachers' Ratings

It will be recalled that the teachers rated the children in their classes in both the Fall and the Spring. The next sections describe the relationship of these ratings to the various sociometric measures, creativity scores, and MCPS subscale scores.

Relationship to Sociometric Ratings and Choices. Our expectations were that teachers' ratings and sociometric nominations would not be related at the beginning of the school year but that some (but not many) significant relationships would be found when Fall sociometric nominations and Spring teacher ratings were considered. We also expected that when Spring sociometric nominations were compared to Spring teacher ratings, many more significant relationships would be found. These, it was expected, would be in the direction of the independent children (rated as such) tending to be chosen or nominated as friends, popular, etc.; and the reverse--non-independent children would tend not to be so nominated.

Table 22 presents the relationship (t-tests) between Fall and Spring teacher ratings to Fall and Spring sociometric ratings (like-dislike) while Table 23 presents the results of chi square analyses of the association between Fall and Spring teacher ratings to items (1) and (5) of the sociometric choice measure for both the Fall and the Spring.

Note: for the Fall comparisons in both tables it was possible to employ the three behavioral (teacher rating) categories of Independence, Dependence, and Pseudoindependence. In the Fall, seven cases were rated as pseudoindependent, permitting the retaining of this as a separate category. In the Spring, however, only two cases were rated

as pseudoindependent, precluding such treatment of data. As a result, in the Spring, teacher ratings were dichotomized into independence vs. non-independence (the latter including both dependent and pseudoindependent cases).

An examination of Table 22 indicates some significant findings: Fall teacher ratings, particularly the dichotomous breaks between independence vs. non-independence and independence vs. pseudoindependence, differentiate significantly, for both the Fall and Spring sociometric ratings, those who receive higher scores (in the expected direction). That is, children who are more liked both in the Spring and in the Fall tended to be rated by their teachers in the Fall as more independent (as opposed to dependent or pseudoindependent). On the other hand, teacher ratings obtained in the Spring did not significantly differentiate any of the sociometric ratings, either in the Fall or the Spring; relationships tended to be in the expected direction, however.

Thus, at least for the sociometric ratings, our overall expectations were not fulfilled. We did find, however, that Fall teacher ratings (but not Spring teacher ratings) were significantly able to predict, in the expected direction, a child's popularity as determined by the like-dislike measures both in the Fall and the Spring.

Table 23 shows that Fall teacher ratings and Fall sociometric choice scores for item (1) are not related nor are Spring teacher ratings and Fall sociometric choice scores for this item. However, several significant relationships emerge with regard to Spring sociometric choice and various teacher ratings: Fall teacher ratings and Spring teacher ratings (particularly when the dichotomy is independence vs. dependence or non-independence) are significantly associated with choice

as measured in item (1) in the expected direction. The picture changes somewhat when item (5) is considered. Fall independent vs. dependent teacher ratings tend to discriminate sociometric choice in the Fall (item 5) in the expected direction and Fall teacher ratings (independence vs. nonindependence or dependence) and Spring teacher ratings significantly discriminate Spring (item 5) sociometric choices in the expected direction.

Thus, in part, our expectations are fulfilled when sociometric choice is considered along with sociometric ratings. Spring teacher ratings and Spring sociometric choice for items (1) and (5) are significantly associated with each other. In addition, however, Fall teacher ratings also significantly predict Spring sociometric choice nominations in the expected direction.

Teachers, it can be concluded, seem to do very well in regard to predicting sociometric status, at least insofar as they are required to rate their pupils as independent, pseudoindependent, or dependent.

Relationship to Measures of Creativity. Our expectations were that various measures of creativity will bear little or no relationship to teachers' ratings at the beginning of the school year, but that in the Spring they will bear some relationship to some of the creativity scores in that they will be positively related to designations of independence, but negatively related to designations of independence and pseudoindependence.

Table 24 presents results of analyses relating Fall teacher ratings to various creativity scores, while Table 25 presents findings with regard to Spring teacher ratings and various creativity scores. These findings do not indicate, as hypothesized, differences between Fall and

Spring teacher ratings in terms of efficiency in predicting creativity scores. However, as Table 24 indicates, one significant finding with regard to Fall teacher ratings did emerge: male children rated as Independent by the teachers in the Fall obtained significantly higher Figural Originality scores than did those male children rated as non-independent. (Note: the latter category contained children rated as both dependent and pseudoindependent.) Table 25 also indicates only one interesting relationship regarding Spring teacher ratings: it is the same relationship as just noted except that the level of significance is not as good. Note: for female subjects, the direction of relationship is the same for this particular creativity score for both Fall and Spring ratings, although a significant level of relationship was not found.

By and large, therefore, teachers ratings were not associated with creativity scores, nor (as predicted)did their Spring ratings show greater validity in predicting such relationships. However, both in the Fall and in the Spring, teachers' ratings of independence for male children tended to be reliably related to scores on the Figural Originality subtest in the expected direction.

Relationship to Missouri Childrens Picture Series Subscales.

Table 26 presents findings in connection with the relationship of high and low scores on each of the MCPS subscales and Fall and Spring teachers' ratings of independence or non-independence. It was expected that fall ratings would bear no relationship to the MCPS scores but that some relationships would be found between teachers' Spring ratings and MCPS scores. The ten 2 x 2 chi square analyses computed (Yates corrected) testing the association of Fall and Spring ratings to each

of the five MCPS subscales yielded no significant findings, as Table 26 discloses. Our expectations concerning these relationships were therefore not confirmed.

"Filler" Status and Ratings of Independence

The Ss in this study were classified according to "filler" status depending on the number of years they were members of the Institute's special enrichment program. We were interested in exploring the possibility that length of exposure to the demonstration program would be associated with status (independence, pseudoindependence, or dependence) as determined through four essential sources: Z-score designations based on the Checklist observations; "global" ratings of Checklist material; and teachers' ratings in the Fall and the Spring (note--all four sources were independently obtained). Table 27 presents frequency distributions of these ratings according to "filler" status, while Table 28 presents results of analyses pertaining to this question. All analyses yielded nonsignificant results, it should be noted.

In the foregoing analyses, to compare extremes in terms of exposure to our program, the Ss were divided into two groups. One group comprised those Ss who received maximal exposure to the program and who had entered either in prekindergarten (1965), designated as E, or in kindergarten (1967), designated as FK. On the other hand, those Ss who received minimal exposure to the program entered in the fall of 1969 at the beginning of third grade and are designated as F3. There were only three F2 subjects (children who entered at the beginning of second grade in 1968) in the current sample. These children were eliminated from the present analyses to permit a critical test to be run on the effects of exposure to the demonstration program. As already noted, "filler" status does not seem to be significantly related to various designations of independence, pseudoindependence, or dependence.

C. Sociometric Measures

Relationship of Ratings--Like-Dislike--(Fall and Spring)--to MCPS Scores. Although no specific expectations were held for results with regard to comparison of the foregoing data, it was felt to be of sufficient interest to proceed with such analyses, nevertheless. It will be recalled (see earlier sections of this chapter) that for each of the MCPS scales, the median was computed separately for boys and girls. Ss at the median were eliminated from further consideration for purposes of these comparisons, and those above the median were designated high, and those below, low, on the particular scale involved. With the two sexes combined, t-tests were run for Fall sociometric ratings and Spring sociometric ratings between the high and low designations for each of the MCPS scales. Tables 29 and 30 present the results. Note, the sociometric rating scores consist of the ratio of the tally for the number of times a child was consistently (test-retest) chosen as liked, to the number of possible choices (the number of children of the same sex). These tables show only one significant finding and one striking trend. For the Fall results (Table 29), it was found that the more aggressive children were more popular ($p < .10 > .05$). In the Spring, however, the more conforming children were found to be more popular ($p < .05$). These findings indicate that the aggressive, attention-getting child somehow makes his presence felt and is chosen as popular by more peers than other types of children, at the beginning of the school year. The picture changes, however, as familiarity increases, and conformity is regarded as apparently more likable--a learning experience that apparently relates to how a child "bucks" the group or molds into it. Further, it may be possible that as a child accepts a group process (the

classroom, the teaching methods, and the like) he begins to like those children who conform to, rather than oppose, that process.

Relationship of Choice (Fall and Spring) to MCPS Subscale Scores. Here again, no specific hypotheses were projected concerning the relationship of sociometric nominations and MCPS subscale scores. Table 31 presents findings in connection with the relationship of scores on Items (1) and (5) for both Fall and Spring administrations of the sociometric choice procedure and each of the MCPS subscales. Only one promising result emerged, and this can be regarded only as a trend, rather than a significant finding: Ss who were classified as high on MCPS Hyperactivity tended to be chosen more often on the basis of Item (5) in the Fall than those children who were regarded as low in Hyperactivity. That the more hyperactive children tended to be more popular (in the Fall but not in the Spring), fits in with a trend reported above regarding the like-dislike scale, in that the more aggressive children tended to be regarded as more popular in the Fall, but not in the Spring. In the Spring, the more conforming children were regarded as more popular.

Relationship of Ratings--Like-Dislike--(Fall and Spring) to Creativity. Table 32 presents Pearson correlation coefficients for Fall and Spring sociometric ratings to the various creativity scores. None of these correlations differs significantly from zero. However, it should be noted that there is a tendency for sociometric ratings and creativity scores to be negatively related in the Spring. That is, there is a trend for the less popular children to score higher on various creativity scores in the Spring, but less so in the Fall. No particular hypotheses were presented with regard to the relationship

of sociometric ratings and creativity scores.

Relationship of Choice (Fall and Spring) to Creativity. Table 33 presents point-biserial correlations and t-tests of significance between Spring and Fall sociometric choice nominations (Items 1 and 5) and the various creativity scores. For the point-biserial computations, dichotomous groups (high vs. low in popularity) were drawn for Items (1) and (5) for both the Fall and Spring testing periods; creativity scores were continuous. For the Figural Originality measure, computations were performed separately for each sex.

Table 33 indicates that in the Fall, Item (1) and the Figural Elaboration score are significantly and positively related to each other. That is, those children who were more popular, chosen more often, scored significantly higher on this creativity subscale (mean score of 16.34) than those children who were less popular--never chosen--(mean score of 10.17). Interestingly, in the Spring, a sharp reversal of this directionality was found. For both Items (1) and (5), there is a significant, negative relationship to the Figural Flexibility score. That is, more popular children were significantly lower on this creativity subtest than other children--children who were never nominated. The more popular group on Item (1) received a mean Figural Flexibility score of 8.57, while the less popular group received a mean score of 9.45. The comparable means for Item (5) were 8.31 and 9.44.

D. Creativity Scores and Missouri Childrens Picture Series Subtest Scores

No particular expectations or hypotheses were presented with regard to the relationship between the variables discussed in this section.

To assess the relationship between status based on MCPS subscale distributions and creativity measures, t-tests were computed. Criterion groups, high and low on each of the MCPS dimensions were compared on each of the creativity measures. These comparisons are reported in Table 34. It should be noted that for the Figural Originality measure, criterion groups were further divided on the basis of sex.

As can be observed in this table, there is a relationship between MCPS Aggressivity and the Verbal Flexibility score ($t=1.73$, $p < .10 > .05$, two-tailed value). Ss low in Aggressivity scores scored higher on the Flexibility measure than Ss high in Aggressivity.

Other significant findings emerged for the criterion groups based on the Hyperactivity measure. Ss high on the MCPS Hyperactivity scale scored significantly higher on the Verbal Originality measure than those Ss who scored low on the MCPS Hyperactivity scale ($t=2.78$, $p < .05$, two-tailed value). Further, those Ss high in hyperactivity level also tended to score higher, though not significantly, in Figural Elaboration ($t=1.69$, $p < .10 > .05$, two-tailed value).

Chapter 5

Summary of Main Findings and Brief Overview

The present chapter summarizes main findings and specifies some directions for future research. Some implications and additional other aspects concerning these findings are also indicated.

A. Scores and Reliability

(1) Revised Classroom Behavior Checklist. Classifications of the children on the basis of Z-score designations were successful in the second year's work, as they were in the first, with little or no differences between schools in the distribution of such ratings. Interobserver reliability information was employed to further refine the checklist by eliminating certain items which fell below an arbitrarily designated criterion level. Despite the time and effort involved in observing children systematically in the classroom, it is felt that not only is this a feasible procedure, but also a necessary one, at least in initial exploratory phases, when concurrent and construct validity information with regard to key concepts is being investigated.

Qualitative use of checklist observations was found to be a reliable procedure. On the basis of reliability explorations (weighted kappa--reflecting degree of agreement between global ratings of two independent raters), it was concluded that the independent raters were able to demonstrate good agreement in designating Ss in the current sample into the three relevant classifications on the basis of summary checklist data.

(2) Creativity Measures. To index the reliability of the scoring

procedures used to obtain the various measures for each S, inter-scorer reliability coefficients were computed. A high degree of agreement was found for the scoring procedures developed. It should be noted that an examination of findings indicated no differences between the sexes on all scales but one: the exception is the Figural Originality subscale on which males scored significantly higher than females. Additional analyses indicated no school differences for any of the scores for any of the creativity subscales.

(3) Sociometric Measures. To index the consistency of the like-dislike scale for each testing period (Fall and Spring), there were test-retest administrations. Using chi square methods which yielded phi coefficients, we found that the association between choices on the first and second administrations of this sociometric rating test was high, significant, and positive for each separate testing period.

Additional analyses disclosed that there was a very high degree of association between sex of rater and sex of child chosen, it should be noted.

Consistency of sociometric choice was also explored for Fall and Spring test administrations separately. In the Fall, for items (1) and (5) only, consistency of choices was found. In the Spring, only item (1) demonstrated significant consistency from test to retest trials. It is interesting to note that the questions in this situation which involved positive choices (e.g., Which two children would you like to sit next to?) were more stable from test to retest than those items which involved negative choices (e.g., Which two children would you not want to sit next to?). Here again, it was found that there was a high degree of association between sex of respondent and popularity,

as indexed by items (1) and (5) in both the Fall and Spring. Note that for both the rating scale and choice situations, these sex-related findings were taken into account when scores for analyses purposes were developed.

Items (1) and (5), it should also be noted, relate significantly and positively to each other both in the Fall and the Spring. It should be further stated that scores for each of these items were stable across test periods.

The relationship between each sociometric choice item and the sociometric rating score was also investigated for the Fall and Spring testing periods, using the point biserial correlation method. These correlations were all significant and positive.

The other essential point should be mentioned, and that is the special difficulties in this population at this age using a group administration method. A review of the development of our methods would describe the host of problems in administration, scoring, and testee-misunderstanding that resulted even after the group method was discarded in favor of the individual method. We feel that a significant aspect of our work lies in the development of our methods, and are especially eager to call the attention of our readers to the techniques we developed for measuring this particular dimension.

(4) Missouri Children's Pictures Series. It should be noted once again that research with the MCPS has been extensive and reliability as well as validity data have been reported. Items which have been employed in all of the subscales of the MCPS show significant and positive discrimination among known test samples. These subscales have also been investigated in terms of internal consistency of items, ten-day retest

reliability, and six-month retest reliability. We should note here that because sex of the Ss enters into the scoring procedures, subscale scores were held separately for each sex for purposes of current data analysis.

(5) Teachers' Ratings. To assess the stability of teachers' ratings, a contingency coefficient was computed relating Fall and Spring designations for each S. This coefficient is highly significant. When separate analyses for schools were conducted, it was found that teachers shifted more frequently in Public School 79 than in Public School 68.

B. Main Findings

(1) Z-Score and Global Checklist Designations: Relationship to Other Variables. As expected, Z-score designations and Global ratings are highly related to each other. Below are more specific findings in connection with these measures:

(a) Relationship to Teachers' Ratings. Z-score designations tend to be positively associated with teachers' ratings in both the Fall and the Spring, and, in accordance with our expectations, the positive relationship is a more confident one for the Spring teachers ratings. Neither Fall nor Spring teachers' ratings were significantly associated with global checklist ratings. Thus, our hypotheses were confirmed only in part in that at least with regard to Z-score designations teachers' Spring ratings show a greater relationship to objective measures than did the Fall teachers' ratings (presumably because teachers got to know their pupils better as the academic year progressed).

(b) Relationship to Sociometic Ratings and Choices. Z- score classifications and sociometric status as determined by the rating methods were not significantly associated either in the Fall or Spring

testing periods. Z-score designations and sociometric choices also were not significantly associated either in the Fall or the Spring. However, global ratings tended to be associated with Spring sociometric ratings (like-dislike), but not with choices either in the Fall or the Spring. An additional trend was found: children rated as more popular in the Spring tended to be among those globally rated as independent on the basis of the Checklist observations. In terms of our hypotheses (see p. 77), it can be said that our expectations were in part confirmed concerning the more popular sociometric status of independent children in the Spring, rather than in the Fall.

(c) Relationship to Creativity Scores. We had expected that creativity scores would be positively related to various designations of independence but negatively related to designations of non-independence derived from the classroom observation Checklist. Overall, although there was a trend for higher means on various creativity scores to be obtained by the independent Ss (designated by both Z-score as well as global ratings), the two significant findings obtained were not in the expected direction. It can be concluded that our expectations concerning the positive relationship of independence and various creativity scores in this group of children were not met.

(d) Relationship to MCPS Subscale Scores. No significant relationships were found between Z-score and global ratings and any of the MCPS scales. It can be concluded, therefore, that our expectations concerning the positive relationship of independence designations and those scales prejudged to represent "better" or "healthier" self-concepts were not met.

(2) Teachers Ratings: Relationship to Other Variables.

(a) Relationship to Sociometric Ratings and Choices. Our expectations were that teachers' ratings and sociometric nominations would not be related at the beginning of the school year but that some (but not many) significant relationships would be found when Fall sociometric nominations and Spring teachers' ratings were considered. We also expected that when Spring teachers' ratings and Spring sociometric nominations were compared, more significant relationships would be found. Predicted relationships would all be in the direction of the independent children (rated as such) tending to be more popular and the non-independent children, less popular. The following significant findings indicate that teachers' ratings tended to relate quite well to sociometric nominations, although the specific hypotheses concerning the greater number of significant relationships regarding Spring teachers' ratings did not hold up for both sociometric measures. Specifically, children who are more liked both in the Spring and in the Fall (sociometric ratings) tended to be rated by their teachers in the Fall as more independent. On the other hand, teacher ratings obtained in the Spring did not significantly differentiate any of the sociometric ratings either in the Fall or the Spring; relationships tended to be in the expected direction, however. Several significant relationships emerged with regard to Spring sociometric choice and various teacher ratings: both Fall and Spring teacher ratings are significantly associated with one of the Spring sociometric choice measures in the expected direction. For the second measure, Fall teacher ratings tended to discriminate Fall sociometric choice in the expected direction; and both Fall and Spring teachers' ratings significantly discriminated Spring sociometric

choice (the other sociometric item) in the expected direction.

(b) Relationship to Creativity Scores. Teachers' ratings in the Fall and in the Spring were not associated with creativity scores; their Spring ratings, thus, did not show (as predicted) greater validity in predicting such relationships. However, both in the Fall and in the Spring, teachers' ratings of independence for male children tended to be positively and reliably related to scores on the Figural Originality subtest in the expected direction. Our overall expectations with regard to various measures of creativity were thus not confirmed. (These were that various measures of creativity will bear little or no relationship to teachers' ratings at the beginning of the school year, but that in the Spring they would show some relationship to some of the creativity scores in that they would be positively related to designations of independence, but negatively related to designations of independence and pseudoindependence.)

(c) Relationship to Missouri Children's Picture Series Subscales. It was generally expected that Fall ratings would bear no relationship to the MCPS scores but that some relationships would be found between teachers' Spring ratings and these scores. Analyses of data yielded no significant findings. Our expectations concerning these relationships were therefore not confirmed.

(3) "Filler" Status and Ratings of Independence. Length of exposure to the Institute's special enrichment program is not associated with status (independence, pseudoindependence, or dependence) as determined through four essential sources: Z-score designations based on the Checklist observations; "global" ratings of Checklist material; and teachers' ratings in the Fall and Spring.

(4) Sociometric Measures: Relationship to Other Variables.

(a) Relationship of Ratings--Like-Dislike--(Fall and Spring) to MCPS Scores. No specific expectations were held for results with regard to these comparisons. Findings were generally non-significant with the exception of one significant finding and one striking trend: Fall ratings of popularity (like) tended to be for the more aggressive children as determined by the MCPS test; and Spring ratings of popularity (like) tended to be significantly related to higher conformity on the MCPS test.

(b) Relationship of Choice (Fall and Spring) to MCPS Scores. Here again, no specific hypotheses were projected concerning these relationships. Findings were entirely non-significant. One trend emerged, however: children who were classified as high on MCPS Hyperactivity tended to be chosen more often on the basis of Item (5) in the Fall than those children who were regarded as low in Hyperactivity.

(c) Relationship of Ratings--Like-Dislike--(Fall and Spring) to Creativity. No hypotheses were presented regarding these comparisons. Analyses of data yielded no significant relationships. But there was a tendency for sociometric ratings and creativity to be negatively related in the Spring.

(d) Relationship of Choice (Fall and Spring) to Creativity. As noted, no hypotheses were presented regarding these comparisons. Here again, findings were largely non-significant, with the following exceptions: Children who were more popular (on one of the two measures) scored significantly higher on the Figural Elaboration score than those children who were less popular in the Fall sociometric testing period. In the Spring testing period, however, two significant negative

relationships emerged, a sharp reversal from the foregoing: for two of the two sociometric measures, more popular children were significantly lower on the Figural Flexibility scale than those children who were less popular.

(5) Creativity Scores and Missouri Childrens Picture Series Subtest Scores. No expectations were presented for the direction of relationships, if any, between and among these variables. Findings were generally non-significant with the following exceptions: children low in MCPS aggressivity tended to score higher on the Flexibility measure than children high in Aggressivity. Children high on MCPS Hyperactivity scored significantly higher on the Verbal Originality measure than those who scored low on MCPS Hyperactivity. Further, children high in Hyperactivity also tended to score higher in Figural Elaboration.

C. Overview and Conclusions

This report describes the last year of work of a two year study designed to systematically evaluate the construct of independence(in a special ghetto population) and its relationship to other variables. In this process, several reliable measurement techniques emerged, including the Checklist, sociometric methods, measures of creativity, and (see Appendix A) various behavioral situations for assessing independent behavior. The Classroom Behavior Checklist shows excellent potential for use in other investigations, as it yields high independent rater agreement in regard to classifying children into the relevant classifications. Methods for assessing creativity were adopted from existing techniques and, once again, high interviewer reliability was found. After considerable pilot-testing and methodological modification, an individually administered method for assessing sociometric status was developed. Test-retest

reliability for the rating scale is high and for the choice situation high for the positive items. The scoring method is both objective and extremely simple. Interestingly, there was a high degree of association between sex of rater and sex of child chosen. Various sociometric measures seemed stable across testing periods and were highly related to each other. We also worked with the Missouri Childrens' Picture Series and found that the children enjoyed working with this reliable, individually administered test immensely.

Thus, a major result of our work involves offering to the professional community our various measurement techniques for further refinement. As for more substantive findings, we offer the following thoughts:

Independence, as a behavioral construct, continues to require careful definition in the context of the specific exploration in which an investigator is involved. Our own definitions seemed to yield reliable and stable data. For example, Z-score designations and global ratings were highly related; further, teachers could be trained in the concepts involved, for their ratings also could be associated with the more objective data. This specific interesting finding is worthy of further exploration as well as dissemination, even if in this interim fashion. That teachers' ratings and the systematic Checklist ratings are associated with each other implies that teachers do quite well in assessing their children, provided some brief training is employed and that there is some conceptual understanding of the constructs involved. As a matter of fact, one of our hypotheses was more or less confirmed with regard to teachers' ratings: as they got to know their pupils better in the Spring, their ratings conformed more to the Checklist designations than did their Fall ratings.

As for sociometric status of the children, there was some indication that popularity and independence (Checklist designations) are indeed associated with each other. Again, the interesting finding in this connection is that teachers' ratings and various sociometric measures are associated with each other in the expected direction in perhaps even a more striking way than ratings based on the Checklist.

By and large, the construct of independence, designated either through teachers' ratings or Checklist classifications, does not seem to be related to the kinds of creativity measures and personality variables we employed. It may well be that the personality dimensions assayed by the MCPS are not relevant ones, and that the particular types of creativity tasks employed also are not relevant to the task we have set out for ourselves.

On the other hand, one particular finding deserves special note and perhaps further empirical investigation: teachers' ratings of independence and Figural Originality creativity scores for males are positively associated.

Sociometric nominations and some of the MCPS data bear interesting associations to each other, again, indicating that further exploration of some of these variables might yield fruitful results. One special question should be raised here, for it concerns the process of change, adjustment, and attitude modification that takes place in the group classroom situation as the academic year progresses. There is some indication that the more popular children are the more aggressive children, determined by Fall sociometric nominations; there is also a trend for the more popular children (in the Fall) to be higher on Hyperactivity. The picture changes in the Spring, however, when the more popular

children tend to be the more conforming children.

Sociometric nominations and some of the creativity scores data also bear interesting associations to each other: there is a suggestion that there is a positive relationship between Fall popularity and creativity, but that once again, in the Spring, popularity and creativity bear a negative relationship.

Does this raise some interesting issues concerning the adjustment and group integration of children in the classroom? Further research can shed some answers.

It should also be noted that some indications of interesting relationships between MCPS Hyperactivity and some of the creativity measures suggest once more some theoretical possibilities connecting, in this age group, movement, curiosity, restlessness, and non-passivity with what is regarded as the creative process. Certainly, this possibility should not be ignored; classroom conformity and passivity--sitting still and passively listening and attending--in the traditional sense, may actually militate against creativity, or at least penalize those children who find such a passive stance difficult.

Finally, it should be mentioned here that these years of effort are being employed currently in the context of another investigation (1970-1971) which is attempting to assay the role of all of the variables we developed in predicting various achievement criteria determined longitudinally or at a given point in time, for this population. This current funded research, involving the same principal investigators, is a logical outcome, hopefully bearing fruit, of the efforts described in the current report.

References

Abel, H., & Sahinkaya, R. Emergence of sex and race friendship preferences. Child Development, 1962, 33, 939-943.

Asch, S. E. Effects of group pressure upon the modification and distortion of judgments. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), Readings in social psychology. (2nd ed.) New York: Holt, 1952.

Asch, S. E. Studies of independence and conformity: I. a minority of one against a unanimous majority. Psychological Monographs, 1956, 70, No. 9 (Whole No. 416).

Baker, E. A. An actuarial study of maturity level in young boys. Paper presented at meeting of the Eastern Psychological Association, Philadelphia, 1969.

Barron, F. Some personality correlates of independence of judgment. Journal of Personality, 1953, 21, 287-297.

Baumrind, D., & Black, A. E. Socialization practices associated with dimensions of competence in preschool boys and girls. Child Development, 1967, 38, 291-327.

Beller, E. K. Dependency and independence in young children. Journal of Genetic Psychology, 1955, 67, 23-35.

Bennett, V. D. C. Development of a self concept Q sort for use with elementary age children. Journal of School Psychology, 1964, 3, 19-25.

Bentley, J. D. The creative thinking abilities and different kinds of achievement. Master's thesis, University of Minnesota, 1961.

Biehler, R. F. Companion choice behavior in the kindergarten. Child Development, 1954, 25, 45-50.

Bonney, M. E. Relationships between social success, family size, socio-economic background, and intelligence among grades III to V. Sociometry, 1944, 7, 26-39.

Butcher, H. J. The personality pattern of creative artists. British Journal of Educational Psychology, 1967, 37, 292-299.

Cartwright, D., & Zander, D. Group dynamics. Evanston: Row, Peterson, 1960.

Cervenka, E. J. The measurement of bilingualism and bicultural socialization of the child in the school setting: the development of instruments. Final report on Headstart Evaluation and Research 1966-67 (contract #66-1).

Cohen, J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, 1968, 70, 213-220.

Crandall, V. J., Preston, A., & Rabson, A. Maternal reactions and the development of independence and achievement behavior in young children. Child Development, 1960, 31, 243-251.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Crowne, D. P., & Stephens, M. W. Self acceptance and self evaluation behavior: a critique of methodology. Psychological Bulletin, 1961, 58, 104-121.

Crutchfield, R. Conformity and creative thinking. In H. Gruber, G. Terrell, & M. Wertheimer (Eds.), Contemporary approaches to creative thinking. New York: Atherton Press, 1962, 120-140.

Davidson, H. H., & Greenberg, J. W. Traits of school achievers from a deprived background. New York: Simon and Schuster, 1967.

Deutsch, M., & Solomon, L. Reactions to evaluations by others as influenced by self-evaluations. Sociometry, 1959, 22, 93-112.

Dunnington, M. J. Investigation of areas of sociometric status groups in a nursery school. Child Development, 1957, 28, 93-103. (a)

Dunnington, M. J. Behavioral differences of sociometric status groups in a nursery school. Child Development, 1957, 28, 103-111. (b)

Edwards, M. E., & Tyler, L. Intelligence, creativity and achievement in a nonselective public junior high school. Journal of Educational Psychology, 1965, 56, 96-99.

Eisenman, R., Platt, J. J., & Darbes, A. Creativity, intelligence, and achievement. Psychological Reports, 1968, 22, 749-754.

Finney, J. C. Some maternal influences on children's personality and character. Genetic Psychology Monographs, 1961, 63, 199-278.

Fishbein, M. Prediction of interpersonal preferences and group member satisfaction from estimated attitudes. Journal of Personality and Social Psychology, 1965, 1, 663-667.

Flescher, I. Anxiety and achievement of intellectually gifted and creatively gifted children. Journal of Psychology, 1963, 56, 251-268.

Fuller, E. M., & Baune, H. B. Injury-proneness and adjustment in a second grade: a sociometric study. Sociometry, 1951, 14, 210-225.

Getzels, J., & Jackson, P. Creativity and intelligence. New York: John Wiley & Sons, Inc., 1962.

Gewirtz, J. L. A factor analysis of some attention-seeking behavior in young children. Child Development, 1956, 27, 17-36.

Grossman, B., & Wrighter, J. The relationship between selection-rejection, and intelligence, social status, and personality among sixth-grade children. Sociometry, 1948, 11, 346-355.

Hartshorne, H. M., May, A., & Maller, J. B. Studies in the nature of character. Vol. 2. Studies in service and self-control. New York: Macmillan, 1929.

Hartup, W. W. Dependence and independence. In H. W. Stevenson (Ed.), Child psychology. The 62nd yearbook of the National Society for the Study of Education. Part 1. Chicago: The National Society for the Study of Education, 1963.

Hasan, P., & Butcher, H. J. Creativity and intelligence: a partial replication of Getzels' and Jackson's study. British Journal of Psychology, 1967, 57, 129-135.

Holland, J. Creative and academic performance among talented adolescents. Journal of Educational Psychology, 1961, 52, 136-147.

Horowitz, E. L. The development of attitude toward the Negro. Archives of Psychology, 1936, No. 194.

Horowitz, F. D. The relationship of anxiety, self-concept, and sociometric status among fourth, fifth, and sixth grade children. Journal of Abnormal and Social Psychology, 1962, 65, 212-214.

Iscoe, I., & Pierce-Jones, J. Divergent thinking, age, and intelligence in white and Negro children. Child Development, 1964, 35, 785-799.

Kanous, E., Daugherty, R. A., & Cohn, T. S. Relation between heterosexual friendship choices and socioeconomic level. Child Development, 1962, 33, 251-255.

Karsten, M. O. The relationship of tested creative abilities and selected factors of academic achievement, intelligence, sex, socioeconomic status, and pupil attitudes. Dissertation Abstracts, 1968, 28 (A), 2557-2558.

Kogan, N., & Tagiuri, R. Interpersonal preference and cognitive organization. Journal of Abnormal and Social Psychology, 1958, 56, 113-116.

Kurtzman, K. A. A study of school attitudes, peer acceptance, and personality of creative adolescents. Exceptional Children, 1967, 34, 157-162.

L'Abate, L., & Hosford, P. M. Screening children with learning disabilities. Paper read at the Invitational Conference on Measurement in Education, Atlanta, October, 1967.

Laughlin, P. R. Incidental concept formation as a function of creativity and intelligence. Journal of Personality and Social Psychology, 1967, 5, 115-119.

Laughlin, P. R., Doherty, M. A., & Dunn, R. F. Intentional and incidental concept formation as a function of motivation, creativity, intelligence, and sex. Journal of Personality and Social Psychology, 1968, 8, 401-409.

Lederman, S. A study of the relationship between parents' self-described dependency attitudes, expectation for achievement in their children and observed dependency behavior in their children. Dissertation Abstracts, 1964, 25(6), 3676.

Lewin, K. Field theory in social science. New York: Harper, 1951.

Lindzey, G., & Byrne, D. Measurement of social choice and interpersonal attractiveness. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology. Vol. 2. Research methods. (2nd ed.) Reading, Mass.: Addison-Wesley, 1968.

Lindzey, G., & Urdan, J. A. Personality and social choice. Sociometry, 1954, 17, 47-63.

Linson, N., & Mussen, P. H. The relation of ego control to overt aggression and dependency. Journal of Abnormal and Social Psychology, 1957, 55, 66-71.

McCandless, B. R., Bilous, C. B., & Bennett, H. L. Peer popularity and dependence on adults in preschool-age socialization. Child Development, 1961, 32, 511-518.

McCandless, B. R., Castaneda, A., & Palermo, D. S. Anxiety in children and social status. Child Development, 1956, 27, 385-391.

McCandless, B. R., & Marshall, H. R. A picture sociometric technique for preschool children and its relation to teacher judgments of friendship. Child Development, 1957, 28, 139-147.

McCord, W., McCord, J., & Verden, P. Familial and behavioral correlates of dependency in male children. Child Development, 1962, 33, 313-326.

Madaus, G. Divergent thinking and intelligence: another look at a controversial question. Journal of Educational Measurement, 1967, 4, 227-235.

Mann, J. H. The influence of racial prejudice on sociometric choices and perceptions. Sociometry, 1958, 21, 150-158.

Marshall, H. R., & McCandless, B. R. Relationships between dependence on adults and social acceptance by peers. Child Development, 1957, 28, 413-419.

Medinnus, G. R. Q-sort descriptions of five-year-old children by their parents. Child Development, 1961, 32, 473-489.

Meer, B., & Stein, M. I. Measures of intelligence and creativity. Journal of Psychology, 1955, 39, 117-126.

Moore, J. E. Antecedents of dependency and autonomy in young children. Dissertation Abstracts, 1965, 26(3), 1766.

Moore, S., & Updegraff, R. Sociometric status of preschool children related to age, sex, nurturance-giving, and dependency. Child Development, 1964, 35, 519-524.

Moreno, J. L. Who shall survive? Washington, D. C.: Nervous and Mental Diseases Publishing Co., 1934.

Moreno, J. L. Sociometry in action. Sociometry, 1942, 5, 298-315.

Newcomb, T. M. The acquaintance process. New York: Holt, Rinehart and Winston, 1961.

Newcomb, T. M. Stabilities underlying changes in interpersonal attraction. Journal of Abnormal and Social Psychology, 1963, 66, 376-386.

Northway, M. L., & Wigdor, B. T. Rorschach patterns related to the sociometric status of school children. Sociometry, 1947, 10, 186-199.

Nunnally, J. C. Tests and measurements: assessment and prediction. New York: McGraw Hill, 1959.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana: University of Illinois Press, 1957.

Owen, D. R. Heritability of personality in children. Unpublished master's thesis, University of Missouri, 1968.

Palm, H. J. An analysis of test-score differences between highly creative and high Millers Analogies members of the Summer Guidance Institute. Master's thesis, University of Minnesota, 1959.

Payne, D. A., & Farquhar, W. W. The dimensions of an objective measure of academic self-concept. Journal of Educational Psychology, 1962, 53, 187-192.

Petersen, R. J., Komorita, S. S., & Quag, H. C. Determinants of sociometric choices. Journal of Social Psychology, 1964, 62, 65-75.

Quimby, V. Differences in the self-ideal relationship of an achieved group and an underachieved group. California Journal of Education Research, 1967, 18, 23-31.

Reese, H. W. Relationship between self-acceptance and sociometric choices. Journal of Abnormal Social Psychology, 1961, 62, 472-474.

Reid, J. B., King, F. J., & Wickwire, P. Cognitive and other personality characteristics of creative children. Psychological Reports, 1959, 5, 729-737.

Rogers, C. R., & Dymond, R. F. (Eds.) Psychotherapy and personality change. Chicago: University of Chicago Press, 1954.

Sines, J. O., Pauker, J. D., & Sines, L. K. The Missouri Children's Picture Series. Unpublished test, University of Missouri Medical Center, 1967.

Sines, J. O., Pauker, J. D., & Sines, L. K. The Missouri Children's Picture Series: an objective, non-verbal personality test for children. Progress Report, University of Missouri Medical Center, 1968.

Sines, J. O., Pauker, J. D., Sines, L. K., & Owen, D. R. Identification of clinically relevant dimensions of children's behavior. Journal of Consulting and Clinical Psychology, 1969, 3, 728-734.

Stein, M. A transactional approach to creativity. In C. W. Taylor & F. Barron (Eds.), Scientific creativity: its recognition and development. New York: John Wiley & Sons, Inc., 1963. Pp.217-227.

Stephenson, W. The study of behavior Q-technique and its methodology. Chicago: University of Chicago Press, 1953.

Taft, R. Predicting school examination results by tests of intelligence and creativity. Australian Journal of Education, 1967, 2, 126-133.

Tagiuri, R. Relational analysis: an extension of sociometric method with emphasis upon social perception. Sociometry, 1952, 15, 91-104.

Torrance, E. Educational achievement of the highly intelligent and highly creative: eight partial replications of the Getzels-Jackson study. Research Memorandum BER-60-18, Bureau of Educational Research, University of Minnesota, September 1960.

Torrance, E. Explorations in creative thinking in the early school years: a progress report. In C. W. Taylor, & F. Barron (Eds.), Scientific creativity: its recognition and development. New York: John Wiley & Sons, Inc., 1963. Pp. 173-184.

Torrance, E. Rewarding creative behavior. Englewood Cliffs: Prentice Hall, Inc., 1965.

Torrance, E. Torrance Tests of Creative Thinking. Princeton, New Jersey: Personnel Press, Inc., 1966.

Trent, R. D. The relationship of anxiety to popularity and rejection among institutionalized delinquent boys. Child Development, 1957, 28, 379-384.

Turner, R. H., & Vanderlippe, R. H. Self-ideal congruence as an index of adjustment. Journal of Abnormal and Social Psychology, 1958, 57, 202-206.

Vernon, P. E. Psychological studies of creativity. Journal of Child Psychology and Psychiatry, 1967, 8, 153-164.

Walker, R. N. A scale for parents' ratings: some ipsative and normative correlations. Genetic Psychology Monographs. 1968, 77, 95-133.

Wallach, M., & Kogan, N. Modes of thinking in young children. New York: Holt, Rhinehart, & Winston, Inc., 1965.

Ward, W. Creativity in young children. Child Development, 1968, 39, 737-755.

Watson, G. Some personality differences in children related to strict or permissive parental discipline. Journal of Psychology, 1957, 44, 227-249.

Winder, C. L., & Rau, L. Parental attitudes associated with social deviance in preadolescent boys. Journal of Abnormal and Social Psychology, 1962, 64, 418-424.

Wittenborn, J. R. Contributions and current status of Q methodology. Psychological Bulletin, 1961, 58, 131-142.

Yamamoto, K. Creativity and sociometric choice among adolescents. Journal of Social Psychology, 1964, 64, 249-261.

Yamamoto, K., Lembright, M., & Corrigan, A. Intelligence, creative thinking, and sociometric choice among fifth-grade children. Journal of Experimental Education, 1966, 34, 83-89.

Table 1

Mean Age and Sex for Institute Third-Graders in Public Schools 68 and 79 Classified According to Years of Exposure to the Enrichment Program

Entered IDS Classes	N	Sex	Classification[a]	Mean Age[b] (Months)
		Public School 68		
Pre-K, 1965-1966 (4 years of exposure)	10	M--4 F--6	E	97.50
Kindergarten, 1966-1967 (3 years of exposure)	0	M--0 F--0	FK	
First Grade, 1967-1968 (2 years of exposure)	0	M--0 F--0	F1	
Second Grade, 1968-1969 (1 year of exposure)	1	M--1 F--0	F2	100.00
Third Grade, 1969-1970 (0 years of exposure)	7	M--4 F--3	F3	102.29
Total, P.S. 68	18	M--9 F--9		99.50
		Public School 79		
Pre-K, 1965-1966 (4 years of exposure)	7	M--4 F--3	E	99.86
Kindergarten, 1966-1967 (3 years of exposure)	2	M--2 F--0	FK	98.50
First Grade, 1967-1968 (2 years of exposure)	0	M--0 F--0	F1	
Second Grade, 1968-1969 (1 year of exposure)	2	M--2 F--0	F2	100.50
Third Grade, 1969-1970 (0 years of exposure)	11	M--4 F--7	F3	104.27
Total, P.S. 79	22	M--12 F--10		102.00
Total, Schools Combined	40	M--21 F--19	17 Es 2 FKs 0 F1s 3 F2s 18 F3s	100.88

[a] F categories contain "filler" Ss, added to the experimental group through the years because of attrition.

[b] As of September, 1969. Mean age for all Ss is just under 8 years, 4 months.

Table 2

Mean Age for Subjects Classified According to Years of Exposure to the Enrichment Program for Both Schools Combined

Classification	N	Mean Age[a] (Months)
E	17	98.47
PK	2	98.50
F1	0	
F2	3	100.33
F3	18	103.50
Total	40	100.88

[a] As of September, 1969.

Table 3

Pearson Correlation Coefficients Indexing Inter-Observer Agreement for Behavior Checklist Items

	Item[a]	r	n[b]
Independence			
4	Helps, sympathizes, shows consideration of others, is thoughtful, offers praise, guidance, assurance	.49	19
5	Sits quietly and attentively when task requires	.63	39
8	Utilizes available free time constructively	.29	40
10	Withstands interference while engaged in individual activities, e.g., doing difficult assignment, a puzzle, painting	.15	34
15	Tries to figure out things for himself before calling on teacher (leader) or other children for help	.10	27
19	Attempts to resolve difficulties that arise between himself & other children without appealing to teacher (leader)	.04	17
Dependence			
1	Withdraws in the face of difficulty	c	4
12	Requires close assistance or direction in order to work at a task	.25	39
14	Lets other children impose on him or boss him around	.26	13
17	Seeks constant recognition and/or approval	.58	39
Pseudoindependence			
2	Is inattentive, is easily distracted by things going on around him	.48	40
3	Uses materials and equipment in a rough or destructive manner	.13	40
6	Brags, shows off, displays exaggerated opinion of own abilities	.13	37
7	Seeks attention, i.e., in a manner that seems designed for active notice rather than stemming from a need for help	.70	39
9	Interferes with the work of others	.46	40
11	Does not follow directions	.47	40
13	Seeks to do things differently from others, even when own method is not effective	c	4
16	Does not pursue task to completion	.18	22
18	Speaks out of turn, interrupts others, takes others' turn	.40	40

[a] Underlined items are retained items after elimination of items with low inter-observer agreement and those with n's that were too small for computing r's.

[b] Number of observations on which the r is based.

[c] Correlation coefficient not computed because n was too small.

Table 4

Checklist Z-Scores and Ranks for Independence, Dependence, and Pseudoindependence Categories By School[a]

ubject ode #	Independence		Dependence		Pseudoindependence	
S 79	Z-score	Rank	Z-score	Rank	Z-score	Rank
666	0.49	28	1.00	33.5	-1.08	6
318	1.27	35	0.50	27.5	-0.16	21.5
289	-0.88	6	0.90	31	-0.10	23
663	1.50	38	-1.38	3	-1.31	4
284	0.31	27	-0.09	20.5	0.71	30
665	-0.52	14.5	-0.39	16.5	-0.16	21.5
667	-0.84	7	0.40	26	-0.27	18.5
271	-2.40	1	1.00	33.5	1.34	37
447	-0.15	22	-0.29	18	-0.50	13.5
661	-0.61	11	-0.09	20.5	-0.27	18.5
449	-0.52	14.5	1.39	37	0.77	31
659	-0.29	20.5	-1.38	3	-0.85	7
288	-0.52	14.5	-1.28	6	0.65	29
277	0.26	26	1.49	38	-0.56	12
312	-1.02	5	-0.39	16.5	1.34	37
272	2.37	40	-1.18	7	-1.20	5
275	-1.34	3	0.90	31	2.09	39
656	-0.79	8	-0.19	19	0.94	33
657	-1.50	2	0.20	24	1.34	37
283	1.36	37	-1.08	8	-0.79	8
664	0.12	24	-0.69	12	-0.44	15
282	1.78	39	-1.38	3	-1.48	3
EAN		19.23		19.75		20.43
S 68						
641	1.00	32	-0.49	15	0.55	27
427	0.77	31	-0.59	14	0.11	24.5
209	-0.29	20.5	-0.99	9	0.28	26
644	1.18	34	1.29	36	-1.54	2
214	-0.74	9	-0.69	12	-0.67	10.5
260	0.68	30	-0.79	10	-0.76	9
234	-0.56	12	1.19	35	1.24	35
233	0.08	23	0.30	25	0.90	32
222	0.17	25	-0.69	12	-0.24	20
638	-0.33	18.5	0.70	29	0.63	28
639	-1.11	4	2.09	40	2.11	40
642	1.04	33	0.00	22	-1.89	1
220	1.32	36	1.99	37	-0.67	10.5
201	0.63	29	0.50	27.5	-0.41	16.5
393	-0.33	18.5	0.10	23	0.11	24.5
224	-0.52	14.5	-1.38	3	-0.50	13.5
212	-0.38	17	0.90	31	1.16	34
228	-0.70	10	-1.38	3	-0.41	16.5
EAN		22.06		21.31		20.58

Note: Rank order is in reverse of usual direction; that is, low rank means lower score on the particular dimension, and less of the attribute presumably being measured.

Table 5

Checklist Rank for Each Behavioral Category and Overall Designation for Each Subject[a]

Subject Code #	Independence	Dependence	Pseudoindependence	Designation[b]
1271	1	33.5	37	NI
2657	2	24	37	NI
1275	3	31	39	NI
2639	4	40	40	NI
1312	5	16.5	37	NI
1289	6	31	23	NI
2667	7	26	18.5	NI
2656	8	19	33	NI
1214	9	12	10.5	M
1228	10	3	16.5	M
2661	11	20.5	18.5	M
1234	12	35	35	NI
2665	14.5	16.5	21.5	M
1449	14.5	37	31	NI
1224	14.5	3	13.5	M
1288	14.5	6	29	NI
1212	17	31	34	NI
2638	18.5	29	28	NI
1393	18.5	23	24.5	NI
1209	20.5	9	26	M
2659	20.5	3	7	I
1447	22	18	13.5	I
1233	23	25	32	M
2664	24	12	15	I
1222	25	12	20	I
1277	26	38	12	M
1284	27	20.5	30	M
2666	28	33.5	6	M
1201	29	27.5	16.5	M
1260	30	10	9	I
1427	31	14	24.5	M
2641	32	15	27	M
2642	33	22	1	I
2644	34	36	2	M
1318	35	27.5	21.5	M
1220	36	37	10.5	M
1283	37	8	8	I
2663	38	3	4	I
1282	39	3	3	I
1272	40	7	5	I

[a]Note: Rank order is in reverse of usual direction; that is, low rank means lower score on the particular dimension, and less of the attribute presumably being measured.

[b]NI = Nonindependent
M = Mixed
I = Independent

Table 6

Pearson Correlation Coefficients (Interscorer Reliabilities for Two Independent Scorers) for Creativity Measures

	Figural Tests		Verbal Tests	
Scores	Picture Completion	Parallel Lines	Product Improvement	Unusual Uses
Fluency	.96	.96	.93	.96
Orginality	.85	.76	.93	.93
Flexibility	.81	.98	.97	.97
Elaboration	.94	.95		

Table 7

Test-Retest Reliability of the Sociometric Ratings: Like-Dislike Scale[a]

Fall

Public School	N[b]	X^2	p[c]	ϕ
68	361	267.82	<.001	.861
79	529	309.07	<.001	.764
Total	890	580.16	<.001	.807

Spring

Public School	N[b]	X^2	p[c]	ϕ
68	306	201.38	<.001	.811
79	420	276.65	<.001	.811
Total	726	481.40	<.001	.814

[a]In these ratings, each child rated every other child as liked or not liked twice in the same session. See text for explanation of the procedure we used to examine test-retest reliability.

[b]N represents the total number of ratings. Each child rated every other child in the class.

[c]Based on 1 degree of freedom for all analyses.

Table 8

Significance of Differences in Popularity Index

(Sociometric Ratings of Like or Dislike)

for Children Rated by Own or Other Sex

Fall

Public School	Sex of Child Chosen	Sex of Rater M	F	t	p
79	M	1.808	1.422	4.834	<.001
	F	1.420	1.749	3.227	<.01
68	M	1.879	1.00	36.016	<.001
	F	1.320	1.802	4.686	<.001

Spring

Public School	Sex of Child Chosen	Sex of Rater M	F	t	p
79	M	1.738	1.432	3.017	<.01
	F	1.391	1.749	2.480	<.05
68	M	1.836	1.00	12.074	<.001
	F	1.498	1.943	3.594	<.01

Table 9

Index of Consistency of Choices
(Test-Retest Reliability) in
Sociometric Choice Situation

Item	Fall		Spring	
	Chi-Square[a]	p[b]	Chi-Square[a]	p[b]
(1) Which two children would you like to sit next to?	18.10	$<.001$	7.67	$<.02$
(2) Which two children would you not want to sit next to?	0.96	$>.10$	0.15	$>.10$
(3) Which two children would you like to play with during lunch, recess, or free time?	0.10	$>.10$	3.06	$>.10$
(4) Which two children would you not want to play with during lunch or free time?	6.85[c]	$<.02$	1.73	$>.10$
(5) Name the two children you like best, like very much.	6.85	$<.02$	0.15	$>.10$
(6) Name the two children you don't like at all; they really bug you.	0.00	$>.10$	0.09	$>.10$

[a]These analyses were computed with one degree of freedom.

[b]All significant findings are in the expected direction (in that a greater proportion of cases was chosen consistently from test to retest) except for the finding indicated by the superscript (c).

[c]These findings are in the unexpected direction in that a smaller proportion of cases was chosen consistently from test to retest.

Table 10

Frequency Distributions of Sociometric Choices for Items (1) and (5)

Item	Number of Choices									N
	0	1	2	3	4	5	6	7	8	
(1)										
Boys										
Fall	6	10	1	2	1	1				21
Spring	10	2	5	3	1					21
Girls										
Fall	10	3	1	1	1			1	1	18
Spring	10	2	2	1	2	1				18
Total										
Fall	16	13	2	3	2	1		1	1	39
Spring	20	4	4	4	6	1				39
(5)										
Boys										
Fall	9	8	1	2	1					21
Spring	11	4	3	2	1					21
Girls										
Fall	10	2	2	1	1				2	18
Spring	11		3	1	1	2				18
Total										
Fall	19	10	3	3	2				2	39
Spring	22	4	6	3	2	2				39

Table 11

Frequency Distribution by Sex of High and Low Scores for the Missouri Children's Picture Series

Subscale	Male	Female
Conformity		
High	7	8
Low	9	7
Maturity		
High	9	9
Low	10	9
Aggressivity		
High	10	5
Low	10	5
Inhibition		
High	10	9
Low	10	9
Hyperactivity		
High	8	7
Low	9	7

Note.--The total possible N for this instrument is 39, 18 females and 21 males.

Table 12

Contingency Data for Teacher's Ratings of S's Reflecting Stability or Change in Ratings

Public School 79

Test Period:	Teacher Rating:	Fall: Independent	Dependent	Pseudo-independent	Can't Rate	Total
Spring	Independent	7	1	2	0	10
	Dependent	0	7	3	0	10
	Pseudoindependent	0	0	1	0	1
	Can't Rate	0	0	0	0	0
	Total Fall	7	8	6	0	21

Public School 68

Spring	Independent	6	0	0	0	6
	Dependent	1	7	0	0	8
	Pseudoindependent	0	0	0	1	1
	Can't Rate	0	1	1	1	3
	Total Fall	7	8	1	2	18

Both Schools

Spring	Independent	13	1	2	0	16
	Dependent	1	14	3	0	18
	Pseudoindependent	0	0	1	1	2
	Can't Rate	0	1	1	1	3
	Total Fall	14	16	7	2	39

Table 13

Frequency Distribution of Teachers' Ratings of Independence for Fall and Spring Testing Periods in Two Schools[a]

Public School	Test Period	Rating			
		Independence	Dependence	Pseudo-independence	Can't Rate
79	Fall	7	9	6	0
	Spring	10	10	1	0
68	Fall	7	8	1	2
	Spring	6	8	1	3
Combined Schools	Fall	14	17	7	2
	Spring	16	18	2	3

[a]The initial N (Fall) on which these data were based was 40. The following is a further breakdown for schools: for Public School 79, the Fall sample consisted of 22 Ss; the Spring sample consisted of 21 Ss. For Public School 68, the Fall and Spring N's were 18.

Table 14

Distribution of Two Independent Raters' "Global" Ratings Based on Summary Checklist Data by Category and Degree

Category	Agreements	Breakdown of Agreements Within Categories According to Degree: High-High	High-Moderate	Moderate-Moderate
Independent	14	7	7	0
Dependent	0			
Pseudo-independent	8	3	3	2
Non-agreements or can't rate[a]	18			
Total N	40			

[a]For 12 Ss, the raters disagreed on category; for 6 Ss there was a can't rate designation. The resulting weighted kappa (K_w) for these data was based on 34 cases.

Table 15

Relationship Among Checklist Z-Score Designations, Global Checklist Ratings, and Teacher's Ratings (Fall and Spring)

Comparisons[a]	N	Phi	Chi Square[b]	p
Teacher Ratings (Fall) with Global Checklist Ratings	20 Total N (40) minus Global Rater disagreements (18) and "Can't Rates" for Teacher Ratings (2)	.02	.010	ns
Teacher Ratings (Spring) with Global Checklist Ratings	20 Total N (39) minus Global Rater disagreements (17) and "Can't Rates" for Teacher Ratings (2)	.14	.375	ns
Teacher Ratings (Fall) with Z-Score designations	22 Total N (40) minus Mixed cases on Z-Score distribution (16) and "Can't Rates" for Teacher Ratings (2)	.36	2.905	.08
Teacher Ratings (Spring) with Z-Score designations	21 Total N (39) minus Mixed cases on Z-Score distribution (15) and "Can't Rates" for Teacher Ratings (3)	.41	3.538	.06
Global Checklist Ratings with Z-Score designations	15 Total N (40) minus Global Rater disagreements (18) and Mixed cases on Z-Score distribution (where Raters agreed) (7)	.74	8.136	$<.01$

[a]These are based on four-fold tables (2 x 2 analyses) that compare independence vs. non-independence on the two dimensions listed.

[b]Yates-corrected values reported.

Table 16

Relationship of Checklist Z-Score Classifications and Global Ratings to Sociometric Ratings (Like-Dislike) in the Fall and Spring

Comparison	N[a]	Mean	t	p
Z-score classifications with Fall Sociometric rating scores	I: 10	.76	.09	ns
	NI: 14	.77		
Z-score classifications with Spring sociometric rating scores	I: 10	.84	.96	ns
	NI: 14	.75		
Global ratings with Fall sociometric rating scores	I: 14	.79	.22	ns
	NI: 8	.80		
Global ratings with Spring sociometric rating scores	I: 14	.88	1.33	$<.10 > .05$
	NI: 8	.75		

[a]Total N (40) minus mixed status Ss on the Z-score distribution resulted in 24 cases, 10 of which were Independent, and 14 Non-independent; total N (40) minus those Ss where there was disagreement between raters in making global ratings resulted in 14 Independent and 8 Non-independent Ss.

Relationship of Checklist Z-Score Classifications and Global Ratings to Sociometric Choice Items (1) and (5) in the Fall and Spring

Comparison	N	Chi Square[a]	p
Z-score classification and high or low status on Item (1) for Fall sociometric choice	(14) Total N of 40 minus 16 cases of "mixed" status on Z-score distributions and 10 unclassified in sociometric choice	.350	ns
Z-score classification and high or low status on Item (1) for Spring sociometric choice	(22) Total N of 39 minus 15 cases of "mixed" status on Z-score distributions and 2 unclassified in sociometric choice	.026	ns
Z-score classification and high or low status on Item (5) for Fall sociometric choice	(16) Total N of 40 minus 16 cases of "mixed" status on Z-score distributions and 8 unclassified in sociometric choice	.000	ns
Z-score classification and high or low status on Item (5) for Spring sociometric choice	(21) Total N of 39 minus 15 cases of "mixed" status on Z-score distributions and 3 unclassified in sociometric choice	.025	ns
Global rating and high or low status on Item (1) for Fall sociometric choice	(16) Total N of 40 minus 18 disagreements for global raters and 6 unclassified in sociometric choice	.485	ns
Global rating and high or low status on Item (1) for Spring sociometric choice	(19) Total N of 39 minus 17 disagreements for global raters and 3 unclassified in sociometric choice	.088	ns
Global rating and high or low status on Item (5) for Fall sociometric choice	(19) Total N of 40 minus 18 disagreements for global raters and 3 unclassified in sociometric choice	.413	ns

(126)

Table 17 (continued)

Comparison	N	Chi Square[a]	p
Global rating and high or low status on Item (5) for Spring sociometric choice	(21) Total N of 39 minus 17 disagreements for global raters and 1 unclassified in sociometric choice	.025	ns

[a]All reported values are Yates-corrected. Four-fold tables with one degree of freedom were employed.

(127)

Table 18

Relationship of Sex and Public School to Creativity Scores

Comparison Group	Verbal Battery Means						Figural Battery Means							
	Fluency	t	Flexibility	t	Originality	t	Fluency	t	Flexibility	t	Elaboration	t	Originality	t
Females[a]	10.28	.54	8.44	.52	3.26	.06	16.11	.25	8.92	.57	11.10	.67	2.36	2.13*
Males[a]	9.25		8.69		3.22		16.43		9.12		13.40		3.02	
P. S. 79[b]	8.79	1.17	8.99	1.67	3.16	.34	16.57	.50	9.00	.19	13.17	.36	2.69	.18
P. S. 68[b]	10.89		8.11		3.34		15.94		9.07		11.36		2.75	

*$p < .05$

[a] N for the entire sample was 39, 21 males and 18 females. One male S was eliminated from analyses of the Verbal Battery because of a zero fluency score.

[b] For the Verbal Battery, Public School 79 provided 20 subjects, and Public School 68, 18. For the Figural Battery, Public School 79 provided 21 subjects, and Public School 68, 18.

(128)

Table 19

Relationship of Z-Score Checklist Designations of Independence (I) and Non-Independence (NI) to Creativity Scores

Creativity Measure	N[a]	Mean	t	p[b]
Verbal Fluency	I---10	11.10	1.75	<.10>.05
	NI---13			
Verbal Flexibility	I---10	7.73	2.10	<.05
	NI---13	9.00		
Verbal Originality	I---10	3.46	.79	ns
	NI---13	2.88		
Figural Fluency	I---10	17.50	1.48	ns
	NI---14	15.14		
Figural Flexibility	I---10	9.11	.14	ns
	NI---14	9.04		
Figural Elaboration	I---10	10.02	.89	ns
	NI---14	14.47		
Figural Originality	Males			
	I----4	3.46	c	
	NI---11	2.75		
	Females			
	I----6	2.85	c	
	NI----3	1.00		

[a] Because one S was eliminated for verbal subtests, the total N for Verbal subtest analyses is 38. For Figural subtest analyses, N is 39. 15 cases were eliminated on the basis of "mixed" status on Z-score classification. The total N for the creativity battery on the basis of the Z-score distribution, then, was 23 for the Verbal and 24 for the Figural subtests.

[b] Two-tailed value.

[c] Males and females were treated separately on this subscale (see Table 18); t-tests were not run because of the small number of N's which resulted.

Table 20

Relationship of Global Ratings of Independence (I) and Non-Independence (NI) to Creativity Scores

Creativity Measure	N[a]	Mean	t	p[b]
Verbal Fluency	I---14	8.93	.36	ns
	NI----7	8.00		
Verbal Flexibility	I---14	8.42	.20	ns
	NI----7	8.56		
Verbal Originality	I---14	3.59	.95	ns
	NI----7	2.85		
Figural Fluency	I---14	16.43	1.38	ns
	NI----8	14.12		
Figural Flexibility	I---14	8.78	2.18	<.05
	NI----8	9.79		
Figural Elaboration	I---14	8.68	1.02	ns
	NI----8	11.17		
Figural Originality	Males			
	I----4	3.14	c	
	NI----5	3.04		
	Females			
	I---10	2.62	c	
	NI----3	1.79		

[a] Because one S was eliminated for Verbal subtests, the total N for Verbal subtest analyses is 38. For Figural subtest analyses, N was 39. 17 cases were eliminated on the basis of Global Rater disagreements. The Total N for the creativity battery on the basis of Global Ratings, then, was 21 for the Verbal subtests, and 22 for Figural subtests.

[b] Two-tailed value.

[c] Males and females were treated separately on this subscale (see Table 18); t-tests were not run because of the small number of N's which resulted.

Relationship Between Independence Checklist Measures (Independence vs. Non-Independence) and Missouri Children's Picture Series Subscale Scores (High vs. Low)

Comparison	N	Chi Square[a]	p
Conformity vs. Z-Score designations	(19) Total N of 39 minus 15 cases of "mixed" status on Z-score distributions, and 5 "Can't Rates" on Conformity	.114	ns
Maturity vs. Z-Score designations	(23) Total N of 39 minus 15 cases of "mixed" status on Z-score distributions, and 1 "Can't Rate" on Maturity	.057	ns
Aggressivity vs. Z-Score designations	(20) Total N of 39 minus 15 cases of "mixed" status on Z-score distributions, and 4 "Can't Rates" on Aggressivity	.449	ns
Inhibition vs. Z-Score designations	(23) Total N of 39 minus 15 cases of "mixed" status on Z-score distributions, and 1 "Can't Rate" on Inhibition	1.046	ns
Hyperactivity vs. Z-Score designations	(20) Total N or 39 minus 15 cases of "mixed" status on Z-score distributions, and 4 "Can't Rates" on Hyperactivity	.008	ns
Conformity vs. Global Ratings	(19) Total N of 39 minus 17 disagreements for global raters, and 3 "Can't Rates" on Conformity	.073	ns
Maturity vs. Global Ratings	(21) Total N of 39 minus 17 disagreements for global raters, and 1 "Can't Rate" on Maturity	.004	ns

(131)

Table 21 (continued)

Comparison	N	Chi Square[a]	p
Aggressivity vs. Global Ratings	(15) Total N of 39 minus 17 disagreements for global raters, and 7 "Can't Rates" on Aggressivity	.033	ns
Inhibition vs. Global Ratings	(22) Total N of 39 minus 17 disagreements for global raters	.042	ns
Hyperactivity vs. Global Ratings	(19) Total N of 39 minus 17 disagreements for global raters, and 3 "Can't Rates" on Hyperactivity	.019	ns

(132)

[a]All reported values are Yates corrected.

Relationship of Fall and Spring Teacher Ratings of Independence, Pseudoindependence, and Dependence to Fall and Spring Sociometric Ratings (Like-Dislike)

N[a]	Teacher Rating[b]	Fall Sociometric Rating Mean	t value	Spring Sociometric Rating Mean	t value
14	Fall Independent	.88	2.09*	.91	1.82**
	vs.				
23	Fall Non-Independent (Dependent and Pseudo-independent combined)	.77		.76	
14	Fall Independent	.88	1.51	.91	1.46
	vs.				
16	Fall Dependent	.79		.79	
14	Fall Independent	.88	2.31*	.91	2.33*
	vs.				
7	Fall Pseudoindependent	.73		.69	
16	Fall Dependent	.79	.89	.79	.78
	vs.				
7	Fall Pseudoindependent	.73		.69	
16	Spring Independent	.87	1.62	.89	1.57
	vs.				
20	Spring Non-Independent (Dependent and Pseudo-independent combined)	.78		.76	

[a]Note that in the Fall, N was 40; however, one S moved out of the district and was not rated by the children on the sociometric instrument; three "Can't Rates" were used by the teachers, resulting in an N of 37. In the Spring, N was 39; however, elimination of "Can't Rate" cases resulted in an N of 36.

[b]Because there were seven ratings of Pseudoindependence employed in the Fall, this category was held separately in further tests with Fall Teacher Rating data. However, in the Spring, Teachers rated only two cases as Pseudoindependent; in this case, it was combined with Dependence, resulting in the Non-Independence category, for all analyses.

*$p < .05$ **$p < .10 > .05$

Relationship Between Fall and Spring Teacher Ratings of Independence, Pseudoindependence, and Dependence and Fall and Spring Sociometric Choice Nominations (Items 1 and 5)

Item 1	Fall Sociometric Choice High vs. Low			Spring Sociometric Choice High vs. Low		
Teacher Ratings[a]	N	Chi Square Value	p	N	Chi Square Value	p
Fall Independent vs. Non-Independent	24	1.891	ns	33	4.482	.05[b]
Fall Independent vs. Dependent	21	2.051	ns	27	5.790	.02[b]
Fall Independent vs. Pseudoindependent	16	.000	ns	17	.165	ns
Fall Dependent vs. Pseudoindependent	11	.006	ns	22	.862	ns
Spring Independent vs. Non-Independent	23	1.483	ns	32	6.036	.02[b]
Item 5						
Fall Independent vs. Non-Independent	27	2.320	ns	33	5.727	.02[b]
Fall Independent vs. Dependent	23	3.812	<.10 >.05[b]	27	5.790	.02[b]
Fall Independent vs. Pseudoindependent	15	.184	ns	17	1.127	ns
Fall Dependent vs. Pseudoindependent	16	1.231	ns	22	.024	ns
Spring Independent vs. Non-Independent	26	4.062	.05[b]	32	5.564	.02[b]

[a]Note that 2 "Can't Rates" were employed in the Teacher Ratings for the Fall, and 3 "Can't Rates" were employed by teachers in the Spring. Further, the reported Ns resulted because of "unclassified" cases for the sociometric choice measure. For the Fall administration of the sociometric choice Item (1), there were 14 such cases; in the Spring, there were 4. For Item (5), in the Fall there were 11 "unclassified" cases; in the Spring, there were 4. Note also that in the Spring, the total N of the sample was 39.

[b]Subjects rated by teachers as Independent were chosen more frequently, and subjects rated as Non-Independent or Dependent were chosen less frequently.

(131)

Table 24

Relationship of Teachers' Fall Ratings of Independence (I), Dependence and Pseudoindependence (NI) to Creativity Scores

Creativity Measure	N[a]	Mean Score	t	p[b]
Verbal Fluency	I--14	11.36	1.22	ns
	NI--22	8.91		
Verbal Flexibility	I--14	8.28	1.28	ns
	NI--22	8.90		
Verbal Originality	I--14	3.43	.70	ns
	NI--22	3.07		
Figural Fluency	I--14	16.07	.50	ns
	NI--23	16.74		
Figural Flexibility	I--14	9.04	.16	ns
	NI--23	8.98		
Figural Elaboration	I--14	11.42	.50	ns
	NI--23	13.27		
Figural Originality	Males[c]			
	I-- 7	3.64	2.52	.05
	NI--13	2.62		
	Females			
	I-- 7	2.83	1.52	ns
	NI--10	2.17		

[a]Total N for these analyses was 37 (39 cases minus 2 "Can't Rates") on teacher ratings in the Fall. Further, because one S was eliminated for Verbal battery, the N for analyses with these subtests was 36; that for the Figural was 37.

[b]Two-tailed values.

[c]Males and females were treated separately for this subscale (see Table 18).

Table 25

Relationship of Teachers' Spring Ratings of Independence (I) and Dependence and Pseudoindependence (NI) to Creativity Scores

Creativity Measure	N[a]	Mean Creativity Score	t	p[b]
Verbal Fluency	I---16	10.56	.56	ns
	NI---19	9.42		
Verbal Flexibility	I---16	8.48	.44	ns
	NI---19	8.70		
Verbal Originality	I---16	3.37	.07	ns
	NI---19	3.33		
Figural Fluency	I---16	16.44	.14	ns
	NI---20	16.25		
Figural Flexibility	I---16	9.09	.68	ns
	NI---20	8.84		
Figural Elaboration	I---16	10.80	.97	ns
	NI---20	14.33		
Figural Originality	Males[c]			
	I--- 8	3.56	2.05	<.10 >.05
	NI---12	2.71		
	Females			
	I--- 8	2.74	1.68	ns
	NI--- 8	2.02		

[a]The total N for these analyses was 36 (39 cases minus three instances of "Can't Rate") ratings by teachers in the Spring. Further, because one S was eliminated for the Verbal battery, the N for Verbal subscales was 35; that for the Figural subscales was 36.

[b]Two-tailed values.

[c]Males and females were treated separately on this subscale (see Table 18).

Relationship Between Teachers' Ratings (Fall and Spring) of Independence vs. Non-Independence and Missouri Childrens' Picture Series Subscale Score (High vs. Low)

Comparison	N	Chi Square[a]	p
Conformity vs. Fall Ratings	(29) Total N of 39 minus 2 "Can't Rates" for teachers, and 8 "Can't Rates" on Conformity	.021	ns
Conformity vs. Spring ratings	(28) Total N of 39 minus 3 "Can't Rates" for teachers, and 8 "Can't Rates" on Conformity	.000	ns
Maturity vs. Fall ratings	(35) Total N of 39 minus 2 "Can't Rates" for teachers, and 2 "Can't Rates" on Maturity	.234	ns
Maturity vs. Spring ratings	(34) Total N of 39 minus 3 "Can't Rates" for teachers, and 2 "Can't Rates" on Maturity	.093	ns
Aggressivity vs. Fall ratings	(29) Total N of 39 minus 2 "Can't Rates" for teachers, and 8 "Can't Rates" on Aggressivity	.021	ns
Aggressivity vs. Spring ratings	(28) Total N of 39 minus 3 "Can't Rates" for teachers, and 8 "Can't Rates" on Aggressivity	.000	ns
Inhibition vs. Fall ratings	(36) Total N of 39 minus 2 "Can't Rates" for teachers, and 1 "Can't Rate" on Inhibition	1.93	ns
Inhibition vs. Spring ratings	(35) Total N of 39 minus 3 "Can't Rates" for teachers, and 1 "Can't Rate" on Inhibition	.288	ns
Hyperactivity vs. Fall ratings	(29) Total N of 39 minus 2 "Can't Rates" for teachers, and 8 "Can't Rates" on Hyperactivity	.066	ns

Table 26 (continued)

Comparison	N	Chi Square[a]	p
Hyperactivity vs. Spring ratings	(28) Total N of 39 minus 3 "Can't Rates" for teachers, and 8 "Can't Rates" on Hyperactivity	.003	ns

[a]All reported values are Yates-corrected.

Table 27

Frequency Distributions of Global Checklist Ratings, Z-Score Checklist Designations, and Teacher Ratings (Fall and Spring) by Length of Exposure to Institute Program ("Filler Status")

Subject Classification	N[a]	Length of Exposure To Institute Program[b]				
		E	FK	F1	F2	F3
(1) Global Checklist Rating	22					
Independent		7	0	0	0	7
Dependent		0	0	0	0	0
Pseudoindependent		4	0	0	1	3
(2) Z-Score Checklist Designation	24					
Independent		4	0	0	1	5
Non-Independent		4	2	0	1	7
(3) Teacher Rating (Fall)	38					
Independent		9	0	0	1	4
Dependent		6	2	0	2	7
Pseudoindependent		1	0	0	0	6
(4) Teacher Rating (Spring)	36					
Independent		9	1	0	1	5
Dependent		7	1	0	2	8
Pseudoindependent		0	0	0	0	2

[a] See Table 28 for explanation of these Ns.

[b] Groups designated as E entered the program at the pre-kindergarten level in September, 1965.
Groups designated as FK entered the program at the kindergarten level in September, 1966.
Groups designated as F1 entered the program at the first grade level in September, 1967.
Groups designated as F2 entered the program at the second grade level in September, 1968.
Groups designated as F3 entered the program at the third grade level in September, 1969.

Table 28

Relationship of Checklist Z-Score Designations, Global Ratings, and Teacher Ratings of Ss (Fall and Spring) to "Filler" Status

Comparison Group[a]	N	Chi Square[b]	p
E and FK Ss versus F3 Ss on Independence and Non-Independence for Checklist Z-scores	22 Total N(40) minus F2 cases(2) and Mixed cases on Z-score designation(16)	.127	ns
E and FK Ss versus F3 Ss on Independence and Non-Independence for Global Ratings	21 Total N(40) minus F2 cases(1) and Rater disagreements (18)	.024	ns
E and FK Ss versus F3 Ss on Independence and Non-Independence for Teacher Ratings (Fall)	35 Total N (40) minus F2 cases (3) and "Can't Rate" ratings on Teacher Ratings (2)	1.612	ns
E and FK Ss versus F3 Ss on Independence and Non-Independence for Teacher Ratings (Spring)	33 Total N(39) minus F2 cases(3) and "Can't Rate" ratings on Teacher Ratings (3)	.857	ns

[a]Those Ss designated E entered the demonstration program in prekindergarten; those Ss designated FK entered in kindergarten; and those designated F3 entered in the third grade.

[b]Yates corrected values reported. All chi squares were based on 2 x 2 contingency tables, with 1 degree of freedom.

Table 29

Fall Sociometric Rating Ratios (Like-Dislike)

by MCPS Designation (High-Low)—(t-Tests)

N	Comparison Groups	Means	t value	p value (2-tailed test)
15	High Conformity	.81	0.229	> .10
	vs.			
16	Low Conformity	.80		
18	High Maturity	.79	0.902	> .10
	vs.			
19	Low Maturity	.83		
15	High Aggressivity	.85	1.721	> .05 < .10
	vs.			
15	Low Aggressivity	.75		
19	High Inhibition	.78	0.990	> .10
	vs.			
19	Low Inhibition	.83		
15	High Hyperactivity	.84	1.587	> .10
	vs.			
16	Low Hyperactivity	.75		

Table 30

Spring Sociometric Rating Ratios (Like-Dislike)

by MCPS Designation (High-Low)--(t-Tests)

N	Comparison Groups	Means	t value	p value (2-tailed test)
15	High Conformity	.91	2.053	<.05
	vs.			
16	Low Conformity	.76		
18	High Maturity	.79	0.404	>.10
	vs.			
19	Low Maturity	.83		
15	High Aggressivity	.86	1.117	>.10
	vs.			
15	Low Agressivity	.77		
19	High Inhibition	.78	0.513	>.10
	vs.			
19	Low Inhibition	.82		
15	High Hyperactivity	.84	0.999	>.10
	vs.			
16	Low Hyperactivity	.75		

Relationship Between Fall and Spring Sociometric Choice Nominations (Items 1 and 5)

and the Missouri Children's Picture Series Subscale Scores (High vs. Low)

	Fall Sociometric Choice				Spring Sociometric Choice			
	N[a]	n[b]	High vs. Low Chi Square[c]	p	N[a]	n[b]	High vs. Low Chi Square[c]	p
Item 1 and MCPS Subscale								
Conformity	22	4	.000	ns	28	7	.000	ns
Maturity	24	2	.078	ns	33	2	1.738	ns
Aggressivity	18	8	.000	ns	26	9	.082	ns
Inhibition	26	0	.008	ns	34	1	.121	ns
Hyperactivity	22	4	2.571	ns	27	8	1.613	ns
Item 5 and MCPS Subscale								
Conformity	24	5	.101	ns	29	6	.284	ns
Maturity	27	2	.717	ns	34	1	.073	ns
Aggressivity	20	9	.314	ns	26	9	.000	ns
Inhibition	29	0	.066	ns	35	0	.325	ns
Hyperactivity	24	5	3.621	$<.10>.05$[d]	28	7	1.555	ns

[a]For Fall and Spring administration of the Sociometric Choice items, the following numbers of unclassified cases resulted, and were subsequently eliminated from analyses: Fall--Item 1--13 unclassified; Item 5--10 unclassified; Spring--Item 1--4 unclassified; Item 5--4 unclassified.

[b]Additional "Can't Rate" designations for MCPS subscales were eliminated from analyses; the n for these cases appear above.

[c]Yates-corrected values are reported.

[d]Children who were classified as high on Hyperactivity were more popular; children low on Hyperactivity tended to be less popular.

Table 32

Pearson Correlation Coefficients[a] for Sociometric Ratings--Like-Dislike--(Fall and Spring) and Creativity Subtest Scores

Creativity Scores[b]	Fall	Spring
Verbal Fluency	.10	.10
Verbal Flexibility	.03	-.12
Verbal Originality	.02	-.13
Figural Fluency	-.10	-.19
Figural Flexibility	.01	-.13
Figural Elaboration	.15	-.15
Figural Originality (Males)	-.15	.06
Figural Originality (Females)	.03	-.29

[a]An r of .32 is required to conclude that the correlation differs significantly ($p=.05$) from zero for both the Verbal and Figural battery. For sexes treated separately, the r for males must exceed .43; that for females must exceed .46.

[b]The N for the Verbal battery was 38. The N for the Figural battery was 39, except for the Figural Originality measure where the 21 males and 18 females were treated separately.

Table 33

Point-Biserial Correlations (and t-tests of Significance) Between Sociometric Choice Scores (Items 1 and 5) and Creativity Subtest Scores[a]

Creativity Scores	Sociometric Choice Testing Period: Fall r_{pb}	Fall t	Spring r_{pb}	Spring t
	Sociometric Item (1)			
Verbal Fluency	.13	1.41	.30	1.80
Verbal Flexibility	.17	.82	-.06	.34
Verbal Originality	.05	.23	.24	1.38
Figural Fluency	-.02	.11	-.17	1.00
Figural Flexibility	-.10	.50	-.41	2.60*
Figural Elaboration	.41	2.21*	-.01	.04
Figural Originality (Males)	.22	.68	.18	.74
Figural Originality (Females)	.19	.70	.05	.39
	Sociometric Item (5)			
Verbal Fluency	.14	.77	.30	1.81
Verbal Flexibility	.20	1.05	-.15	.85
Verbal Originality	.13	.69	.15	.90
Figural Fluency	-.06	.31	-.23	1.33
Figural Flexibility	-.12	.66	-.51	3.40**
Figural Elaboration	.29	1.59	.04	.25
Figural Originality (Males)	.35	1.22	.07	.29
Figural Originality (Females)	.05	.17	.13	.51

[a]For Item (1) in the Fall, the N for all computations was 26 (10 highs and 16 lows), except for Figural Originality, where the eleven males and fifteen females were treated separately. In the Spring, the N on the Verbal subtests was 34 (15 highs and 19 lows); that for Figural subtests was 35 (15 highs and 20 lows), except for the Figural Originality measure where the nineteen males and sixteen females were treated separately. For Item (5), in the Fall, the N for all computations was 29 (10 highs and 19 lows), except for Figural Originality, where the thirteen males and sixteen females were treated separately. In the Spring, the N for the Verbal subtests was 34 (13 highs and 21 lows); that for Figural subtests was 35 (13 highs and 22 lows), except for Figural Originality, where the seventeen males and eighteen females were treated separately.

*$p < .05$ (two-tailed); **$p < .01$ (two-tailed).

(5hD)

Mean Creativity Scores for High or Low Status Groups

on Missouri Children's Picture Series Subscales

Creativity Measure[a]		Status on MCPS Subscales: Conformity High	Conformity Low	t	Maturity High	Maturity Low	t	Aggressivity High	Aggressivity Low	t	Inhibition High	Inhibition Low	t	Hyperactivity High	Hyperactivity Low	t
Verbal Fluency	N	15	15		17	19		15	14		19	18		15	15	
	Mean	10.67	8.40	1.18	10.00	10.00	.00	10.67	7.64	1.64	9.21	9.50	.16	10.20	8.53	.83
Verbal Flexibility	N	15	15		17	19		15	14		19	18		15	15	
	Mean	8.28	8.88	1.13	8.25	8.83	1.19	8.10	9.05	1.73**	8.92	8.33	1.29	8.68	8.38	.58
Verbal Originality	N	15	15		17	19		15	14		19	18		15	15	
	Mean	3.38	3.10	.47	2.88	3.58	1.30	3.29	2.82	.82	3.24	3.18	.13	4.20	2.72	2.78*
Figural Fluency	N	15	16		18	19		15	15		19	19		15	16	
	Mean	15.07	15.19	.62	15.89	16.47	.44	16.47	16.00	.31	16.68	15.79	.70	16.47	17.19	.51
Figural Flexibility	N	15	16		18	19		15	15		19	19		15	16	
	Mean	9.02	8.93	.21	9.11	9.02	.27	9.14	8.88	.72	8.87	9.12	.70	8.87	9.16	.74
Figural Elaboration	N	15	16		18	19		15	15		19	19		15	16	
	Mean	11.40	10.32	.38	10.21	14.55	1.23	14.32	10.71	.84	10.79	14.13	.96	15.53	9.14	1.69
Figural Originality[b]																
Males	N	7	9		9	10		10	10		10	10		8	9	
	Mean	2.52	3.28	1.61	3.11	3.03	.18	3.25	2.71	1.25	2.76	3.15	.92	3.30	2.78	1.13
Females	N	8	7		9	9		5	5		9	9		7	7	
	Mean	2.60	1.84	1.57	2.27	2.46	.41	1.93	2.38	.82	2.48	2.24	.52	2.66	2.45	.40

[a] Because one S was eliminated from the Verbal subtests, Verbal and Figural Ns differ by one.

[b] Males and Females were treated separately for this subscale of the creativity battery.

*p < .05, (two-tailed value); **p < .10 > .05 (two-tailed value).

Appendix A
Final Report for Interim Research Period 9/1/68 to 8/31/69
OEO Contract No. B89-4612A

An Exploration of Transfer of Independent Behavior from Enrichment Classrooms in Harlem to a Non-Classroom Behavioral Situation

Chapter 1

Background and Hypotheses of the Current Investigation

General Background and Need for this Investigation

In 1962, in an effort to reverse the corrosive process of academic retardation that plagues the disadvantaged child, the Institute for Developmental Studies began to translate its prior and current research findings into action in the form of a demonstration program in Harlem public schools, carried out in cooperation with the New York City Board of Education with the support of the Ford Foundation, Office of Economic Opportunity, and other public and private funding agencies. (See Pavledge [1967] for a description of the Institute's program.)

Our enrichment program stemmed from the assumption that heavy stress on cognitive areas--specifically on verbal and language skills--would be the most effective means for intervening into the ways of thinking, learning, and behaving (with the consequent limitations in achievement) displayed by the urban, ghetto children with whom we were working. For many years, we have employed a variety of instruments for assessing the effectiveness of our work in Harlem. These have included such instruments as: The Stanford-Binet Intelligence Scale, the Peabody Picture Vocabulary Test, the Columbia Mental Maturity Scale, Reading Prognosis Test, Lorge-Thorndike Intelligence Test, Gates Reading Test, Illinois Test of Psycholinguistic Ability, the Wechsler Intelligence Scale for Children, and the Institute's specially developed Early Childhood Inventories.

As an ongoing evaluative process, the Institute continually explores various behaviors as measured by the foregoing instruments, but also, in

a more general way, attempts to describe certain qualitative aspects of our program based on such sources of information as supervisors' and teachers' reports, "feedback" from parents and community observers, and principals' comments and observations. In general, findings have been most heartening, since many positive results have emerged.

Of significance and relevance to the purposes of this proposal, we should note that through the years, supervisors, teachers, visitors, and community personnel have been impressed with certain "noncurricular" aspects of our interventive impact on the children in Institute classes. Expressions such as "greater independence," "more maturity," "quieter and less motoric" have been applied to them, frequently in the context of comparisons to children in other ghetto classrooms. And indeed, it would seem that many of the teaching techniques and classroom procedures employed by the Institute are not only directed toward the teaching of <u>content</u> but also toward introducing changes in behavior, such as the developing of independence in the learner. Thus, such classroom techniques as the introduction of individualized activities, quiet work-time, small group involvements, decision-making activities, individual use of tape-recorder, etc. have not only been employed to teach, for example, reading, listening, and language skills, but also to aid in the development of certain behavioral and interpersonal skills. Indeed, the classroom is physically engineered to encourage the children to move about, obtain and then replace materials, and to operate equipment. We have been trying to give them opportunitites to see that they can have substantial control over their environments.

We were soon confronted with the questions of whether the Institute's enrichment classes have indeed encouraged the kinds of behaviors various

observers have described. Even more important, does the behavior displayed in the demonstration classrooms "carry over"--transfer--to situations outside of the classroom?

Some Theoretical and Methodological Issues. It is important to note that several issues are relevant to our research. The first concerns the need for behavioral explorations of this sample for purposes of assessment.

(1) The need for behavioral criteria. It bears repeating, but not necessarily elaboration (since so much as already been written about this issue--see Eells, et al. [1951]; Guidelines for Testing Minority Group Children, Deutsch, et al. [1964]; Minuchin, et al. [1967]; and Zubin, Eron, & Schumer [1965]), that standardized tests, especially of the paper-and pencil variety (but also even individual tests, say, of general ability), are not culture-free. They penalize large numbers of children from disadvantaged social environments because they require skills in test-taking, listening, following instructions, and sitting still, not to mention general verbal and language ability which are simply not "built-into" the repertoire of children who otherwise may be intelligent, alert, and well able to learn. Further, from a more theoretical view, and in connection with what is emerging as methodologically desirable, we are learning that behavior is the most valid and reliable criterion for diverse measurements in personality, psychotherapeutic change, diagnosis, prediction of success in various job situations, etc., even for nondisadvantaged populations.

We note, then, that our orientation rested on the general assumption that employing aspects of behavior to predict behavior and to shed light on behavior appeared to be the most methodologically sound approach.

(2) The need for employing a "transfer model" and problems in interpreting results. Any individual working in the area of test development, with its criterion problems as well as ultimate predictive validity problems, knows (whether or not he gets around to this) that eventually his generalizations or recommendations must suffer the test of cross-validation or the test of applicability to new samples or new situations.

In the same vein, we felt that it was not enough to demonstrate that our classes produced certain types of behavior; we wished to demonstrate that this change carried over to other situations. A child "programed" by certain types of classroom environments may behave in accordance with that programing provided no change in milieu takes place. Crucial for study, not just in the proposed research, but also with regard to teaching techniques in general, was the extremely important question of transfer and generalization to other situations. Indeed, an implicit assumption of interventive education as well as conventional education is that the child generalizes what he learns (or what we think he has learned) to other classrooms, to the community, and to his daily life.

Thus, the long history in academic and laboratory circles that the transfer of training paradigm has enjoyed seemed to us more than a wasted effort in experimental minutiae. For this history offered a model for experimental exploration of a vital issue. Are we indeed making an impact which has meaning outside the teacher-classroom situation?

It should be stressed, however, that this issue is not that clear nor that easily resolved; definitive answers may not be readily available to the researcher. For intrinsic to the problem of transfer is the question of the role of situational variables. Perhaps, after all, independent behavior, at least in this age group, is partially rooted in situational and stimulus-

milieu variables, and that only a careful examination of the surrounding situation (above and beyond individual personality variables) will yield accurate predictors of behavior. One implication of this issue is that failure to find transfer effects may not at all mean that we have failed to train for independence. The behavioral sessions and their tasks are, after all, rather dissimilar from the classroom situation, and the behavior elicited in them. Independence may not be, in other words, a personality trait (as noted, at least at this age level) that is "carried" within the individual from situation to situation.

Positive findings in the current study, then, will suggest transfer effects. Lack of positive findings, however, must perforce be less definitively interpreted.

Objectives and Hypotheses

As noted in the original proposal for this investigation, one of the major purposes of the current research lies in its implications for the further understanding of the independent behavior of children in and out of the classroom; an equally important purpose is relevant to assessing the effectiveness of certain Institute classroom procedures. Specifically with regard to the latter point, the basic problem under consideration is whether noncurricular aspects of behavior transfer to new situations.

It was hoped that the present investigation would also shed light on whether independence is correlated with length of exposure to the IDS enrichment classes, as well as whether there are individual variations in the extent to which independent behavior transfers to unfamiliar situations. We also hoped to ascertain whether non-Institute control Ss from the same schools could be differentiated from the Institute experimental Ss in regard to the extent to which the behaviors related to the dimensions in which we

are interested would emerge in the transfer situation.

Our objectives, then, related to assessing the Institute's effectiveness in encouraging independent, autonomous behavior in its enrichment classrooms, on the general (commonly accepted) assumption that not only is such behavior developmentally and socially "normal" on the age levels with which we were working, but also indicative of the kinds of cognitive and emotional changes that underly productive, achieving future behavior--in and out of the classroom. To evaluate our success along these lines, we used behavioral rather than standardized test criteria, both in the classroom and in a "transfer" situation, to explore whether the behavior that appeared in the classroom did, indeed, emerge outside of that "programed" situation. Moreover, we wished to explore whether, in the transfer situation, relatively greater frequencies of independent behavior would be found among our demonstration pupils than among an equivalent group of disadvantaged controls.

The study, in effect, was geared to evaluate our curriculum effectiveness and goals in terms other than those related to achievement, content skills, and the like. Of significance in the design of this study is the meaningful issue of transfer of learned behavior to situations outside of the classroom with the expectation that some children in our experimental sample would not display independent behavior--a reflection of the fact that an interventive program cannot, and does not have the same impact on all members of its target population, and that there are subgroups among the disadvantaged population itself.

The general model we employed to achieve the foregoing goals was to observe (in the classroom) and then rate third-grade children from the demonstration program on systematically developed, reliable, behavioral checklists.[1]

---

[1] Equivalent control classes are in classroom situations that are conducted conventionally, with the typical classroom arrangements of seated-pupils-facing-a-teacher-at-a-desk layout and with movement, free choice of materials, etc., not encouraged. Such Ss could not, perforce, be rated and compared to the demonstration Ss.

The latter instrument-development procedure proved to be a major aspect of the year's work. Two trained, independent experimenters observed the experimental Ss in the classrooms over a period of time, employing the behavioral checklist. Two new raters then anonymously rated the material based on behavioral observations of the children along a four-point discontinuous scale: independent (productive); pseudoindependent; nonindependent; and can't rate, don't know. (The"pseudoindependent" classification derives from the basic fact that behavior that subsumes motility, detachment, noninvolvement, inattentiveness, and the like--might be rated as independent, but that such behaviors could hardly reflect the construct in which we were interested.) Following a behavioral, non-classroom observed transfer situation which allowed independent behavior to emerge (and the development of transfer tasks also proved to be a major research activity), observers who did not observe in the classroom rated random small groups of demonstration and equivalent control subjects on the same four-point discontinuous scale. No rater had any prior knowledge about any S.

Our hypotheses and expectations based on the above procedures (the latter are described in complete detail in the next chapter) follow:

(1) After operationally defining so-called independent, mature (age-related) behavior, we hoped to be able to distinguish the behavioral variables associated with this attribute from "independent" behavior which is not productive or useful; and these behaviors, in turn, from nonindependent behavior. We expected a range of such behaviors in the enrichment classrooms with a relatively larger number of "independent" ratings than the other types of ratings; and, we also expected that F3 "filler" Ss (replacements due to attrition who entered our program in the third grade), since they had virtually no years of exposure to the enrichment program, would be rated relatively less frequently as "independent" than the other experimental Ss.

(2) Having thus identified behavioral subgroups along this dimension in the demonstration classes, we expected that in a non-classroom, behavioral transfer situation, those rated as "independent" would tend to receive the same rating by another team of observers who had no prior knowledge about the preceding ratings; the same expectations were held for the Ss rated as "pseudoindependent" and "nonindependent." In other words, we expected our prior classification along this trichotomy to be related to classifications of the same Ss as observed and rated in a new situation. "Filler" status would of course be considered in this analysis.

(3) We expected that control Ss would receive fewer ratings of "independent" in the transfer situation, and relatively more ratings of "dependent" than the demonstration Ss, again, with "filler" status of the experimental Ss considered.

(4) Needless to say, underlying the past year's research activities was the basic expectation that we would develop a reliable, usable classroom checklist, which we could later employ in more theoretical explorations of independence along construct validity lines.

(5) A subsidiary expectation concerned the notion that on the basis of checklist designations, trained "global" raters (without seeing the Ss) would rate some of them as "pseudoindependent," and that the latter is a meaningful category in this regard.

Finally, we should add that we were quite interested in exploring the role of situational variables in determining behavior, on an assumption that independence is not, after all, "carried within" the individual from situation to situation regardless of the nature of that situation.

Chapter 2

Sample and Methodology

The general model for the currently described research was briefly presented in the preceeding chapter. The design included: (a) the development of a reliable behavioral classroom checklist for independent, pseudo-independent, and dependent behavior; (b) the formal phase of classroom observations (over time) using the checklist (involving two observers); (c) the "global" rating of these checklists by two different independent raters; (d) the development of quantitative, continuous scores based on checklist observations; (e) the development of behavioral tasks for the "transfer" sessions which would yield an opportunity for reliable ratings to be made; and (f) the "running" of experimental and control Ss in these transfer sessions with independent observations and ratings made by pairs of observer-raters. Note, the first (classroom) phase of the research utilized children (the experimental Ss) from the Institute's classes. Non-Institute classes were not observed for independent behavior, since, in terms of their physical layout, they did not offer an opportunity for relevant ratable behaviors to emerge. The behavioral sessions, on the other hand, were based on small groups of randomly assigned experimental and control children. The Institute children had, it should be noted, varying degrees of exposure to our enrichment classes.

The current chapter describes: (a) the subject population with which we worked in this investigation; (b) development and description of the classroom behavior checklist; (c) development and description of the behavioral tasks for the transfer sessions; and (d) description of the global rating employed in the various phases of our study. Chapter 3 presents information concerning the development of other scores, the checklist revision, various reliability explorations we conducted, and data analyses and findings for various parts of the study.

Sample

Experimental Ss--Phase I (Classroom Observations)

The larger population from which subjects for this study were drawn consisted of third graders (1968-1969) in the Institute's demonstration classes in Public Schools 68, 79, 90, and 175 in Harlem.

Because the currently described study was of only one year's duration, we could not work with third graders from all four of the foregoing schools. Had we done so, about 150 subjects would have been involved when equivalent controls were added. Since this was not a feasible sample size, two schools were selected on the basis of their yielding the largest numbers of third-grade children who have had at least two years of exposure to the Institute's enrichment program.

In the classroom observation phase of the research, 19 children were observed at Public School 90 and 21 children at Public School 175. Of the 40 children included in this sample, 28 children had at least two years of exposure to the Institute's program (12 at Public School 90, and 16 at Public School 175). The distribution of this sample by school, age, sex, and length of time in the program is shown in Table 1.[2]

Table 1 shows that average age for each school is virtually the same, and that only minor age differences for the various types of subjects within each of the schools are found. Most Ss, as of September, 1968, were eight years and a few months old; a few Ss were just under eight years of age; one S was seven years and four months, and another, seven years and five months old. Sex distribution is uneven in this population, with almost two-thirds of the total sample consisting of boys.

Since years of exposure to the IDS program is a variable in which we

---

[2]The first progress report (December, 1968) described a sample of 41 children (see Table 1 in Progress Report #1). One S at Public School 90, however, moved out of that school district before the classroom observations could be completed.

are interested, and since the development of independence is age-related, it is interesting to see whether or not wide differences in mean ages are found between subjects classified as to "filler" status. Table 2,which presents mean age for various classifications of Ss (with both schools combined), indicates that as of September, 1968, average age for each classification is approximately 8 years and a few months, with wide differences among groups not found.

Experimental and Control Ss--Phase II (Transfer Sessions)

The Ss described above were observed along with control Ss in the behavioral transfer sessions which will be described shortly. The actual experimental sample was reduced by one because a female S from Public School 90 was not available when the behavioral transfer sessions were run. She was an F3 S, thus reducing that group from 9 to 8 Ss and the total experimental sample in the transfer sessions from 40 to 39. Table 3 presents the distribution of this sample by school, age, sex, and length of time in the program, and Table 4 presents mean age for these classifications with both schools combined. When Table 4 is compared to Table 2, it can be seen that only minor changes were introduced by the elimination of one S.

Control Ss were drawn from the third-grade classes of Public Schools 90 and 175. No selective factor was introduced: controls were taken from all the non-Institute third-grade classes in these schools. Table 5 presents mean age and sex breakdown for these subjects. Note, although 39 control Ss were actually "run" in the behavioral sessions, only 37 were regarded as control Ss due to sampling errors.

The control Ss come from the same SES, school, and community backgrounds as the experimental Ss. Mean ages are almost identical with those of the experimental Ss, it should be noted.

The Classroom Behavior Checklist

This section describes the development of the Classroom Behavior Checklist for Dependent, Independent, and Pseudoindependent behavior, as well as the checklist itself as we actually used it in initial phases of the current investigation.

Working Definitions of Behavioral Categories

During the early periods of pilot work, when we were considering items for checklist inclusion, we clarified our thinking concerning what we meant by the various behavioral categories. It should be stressed that this clarification took place within the context of constant feedback from our research team as to their observational experiences with groups of children using some of the earlier items. It was during this early period, for example, that we confirmed our expectation that our prior classification, pseudoindependent, had some behavioral meaning. (We saw, however, that this term could be ambiguously interpreted, since it seemingly did not imply a behavioral referent other than that which could be regarded as a residual or derivative from the independent behavior category.) Raters, teachers, supervisors, and others on our professional staff all seemed to agree that certain behaviors as observed in the classroom fell neither into an independent nor a dependent category, but yet were still classifiable--that is, such behaviors were easily identifiable and were not of a "don't know" nature. Indeed, this third category (but not middle category, since a continuum is not implied) subsumed behaviors which our observers suggested were rather salient and recognizable, for their characteristics were overt and often possessed an attention-getting quality.

From sources in the literature, our own experiences in the schools, and feedback from our research team, then, the following working definitions

of the behavioral categories emerged:

(1) Independent behavior is oriented toward productive (constructive) goals, is age-related and is self-initiated. The independent child manifests behavior which is both self-reliant and self-assertive. He is considered independent of others to the extent that he attempts to satisfy his needs in a manner which is not entirely a function of the way others behave toward him. A child is considered independent when he conducts his problem-solving activities without seeking unnecessary help from others, and is able to carry on his activities without continuous need for reassurance, affection, or approval. Independence also includes behavior which is self-assertive, such as striving to master tasks. Independent children characteristically show relatively good motor control, that is, their motility is appropriate to the situation: they are neither overly constricted nor excessively motoric.

(2) Pseudoindependent behavior can be described as nonproductive, nondependent behavior, characterized by the seeking of attention through asocial or antisocial activities, the maintenance of physical and psychological distance from peers or teachers (leaders), high motility, and inconsistency in the pursuit of goals. Children so designated are likely to interfere with the activities of other children, refuse to participate in group activities, and be destructive of work materials and equipment. Such children often do not accept help or follow instructions, even though their own efforts are clearly not successful. They also tend to show random sequences of behavior with unrelated activities begun and discarded before completion.

(3) Dependent behavior is determined by the presence of specific characteristics, not merely by the absence of characteristics specific to independence. Responses that are characterized as dependent, are often those that seem directed toward obtaining social reinforcement. Such behavior includes many different kinds of activity, such as: persistently seeking help (whether necessary or unnecessary); affection-seeking; and bids for

attention, reassurance, approval, succorance, and guidance. Dependent behavior is also manifested in passivity and lack of spontaneity.

Development of the Classroom Observation Checklist

Early in September, 1968, various classroom behavior scales and reports in the literature were examined for the construction of a large pool of possible items. Some of these sources were: Beller (1955); Cervenka (1967); Davidson and Greenberg (1967); Finney (1961); Gewirtz (1956); and Hartup (1963). As noted, this procedure was also combined with group discussions and feedback concerning refinement and clarification of what was meant by each of the categories (independent, dependent, and pseudoindependent). Later that month, each item was discussed from the point of view of a minimal standard--face validity. At this point, it became clear that items reflecting motor activities (such as hyperactivity in the classroom) and distractibility should be included in the pool of items; it also became clear that there were many seemingly distinct items which were, in fact, reflecting identical behavioral elements. Gradually, more items reflecting the child's interaction with other children were developed. We also began to consider, at this time, the mechanics of the actual observations.

Pilot-testing was begun in October, 1968, using second-grade children in Public Schools 175, 90, and 79 in Harlem. Third-grade children at the Friends Seminary were also observed. The original checklist was revised three times on the basis of pilot observations. During this preliminary period, the observers worked with and evaluated the feasibility of using various observation units (that is, size of units or groups of children who were observed), durations of observation periods, and other classroom observation procedures.

In the light of certain difficulties encountered in observing the children in small groups, or in too short segments of time (the behaviors in which we were interested required some continuity of observation-over-

time), as well as because of other mechanical considerations (for example, the unwieldiness of checking the behavior of several children at a time on large checklists), the procedure finally adopted in November was: two observers independently observed each child in a single classroom for five-minute periods, in sessions lasting up to two hours. The particular hours selected were to be scheduled well in advance and took into consideration the teacher's daily schedule of classroom activities. (We learned, for example, that observing children cooking, or having juice, or dancing in a group did not yield the kind of variation in behavior which we were seeking.) A brief "narrative" was written by the observer as he was watching the particular child. At the end of each day of observation, each observer independently reviewed his own narratives, and the relevant behavior was checked on the checklist (for each child) along a four-point scale--frequently, moderately often, rarely, and never. A fifth category, no opportunity to observe, was also included.

The checklist at this point had been revised to include nine items in each of the three behavioral categories (see next section). By about the middle of November, the checklist was tested once more (Public School 175) by having each observer observe and rate each child for a five-minute period. The two observers showed high reliability in their observations.

The formal data collection period began in November, 1968, and ran through most of January, 1969, in Public Schools 175 and 90. Subjects were independently observed by two observers in their normal classroom settings. The total number of sessions and observations per child ranged from fourteen to twenty-one, with a median of twenty. No difficulties in either using the checklist or in the actual mechanics of classroom observations arose during the testing period.

The schedule for both observers was based on random selection of days of the week and random time-cluster sampling at each school (modified only to balance the number of observations at each school). In addition,

separate randomized lists of subjects were prepared for each observer for each day of observation.

Departures from strict randomization of the ordering of subjects were necessitated, however, by the absence of some of the children for various reasons (e.g., some were not in school on a particular day, some were sent on errands, some were taken out to rehearse their parts in a forthcoming play, and sometimes the group activity so structured the situation that valid observations could not be made, as in story-telling time).

Whenever possible, a child who was not available at the scheduled time was observed later on a randomly determined basis. If a child was absent more than four times, the observation period was extended so that the ratings were finally based on a minimum of 16 (with a maximum of 20) observations on each child for each observer.

At the end of each day, all identifying data were removed from the narratives after they were typed. At the end of the several-week observation period, each observer independently summarized his daily checklist on one master list for each child, with frequency counts summarized for each category for the four-point scale: frequently, moderately often, rarely, and never (as well as no opportunity to observe).

Note, narratives allowed the opportunity for singular or unusual events to be recorded--events that were not anticipated, or even though they occurred but once, were thought to bear some significance to what we were after. Narratives also could provide, we thought, an opportunity to qualify a behavior tally on the checklist--for example, a child who was considered unusually hyperactive was later taken home by his mother because the nurse reported that he had a high temperature. This procedure was introduced in the light of the possibility that some unique aspect of behavior or some special circumstance could arise which was not or could not be anticipated by the behavior checklist, despite our careful pilot-testing in many classroom situ-

ations. Such narratives, we expected, would allow the final raters to take into consideration a qualifying, explanatory, or modifying circumstance which might alter the picture obtained only from frequency counts based on the checklist. In actuality, our classroom observers have noted that only in rare instances did such narratives "clarify" the checklist information, and they were of little or no help when the final global ratings were done. Further analysis of the narratives was therefore not undertaken.

The Classroom Observation Checklist Items

Appendix A presents the behavior checklist as employed by two independent observers of all the children in the same classroom situation (at different times during at least a two-hour period at each school) on different days over a period of almost two months. The sequence of items is a random one. Below are the checklist items (with examples of actual behavior) regrouped into categories according to what we think they are reflecting, as judged by construct considerations, as well as empirical considerations during pilot phases of our exploration.

Independence Items--9

(4) Volunteers contributions to class discussions and projects--e.g., offers information without being called upon.

(5) Helps, sympathizes, shows consideration of others, is thoughtful, offers praise, guidance, assurance--e.g., helps other with work when appropriate, lends or shares materials freely.

(6) Sits quietly and attentively when task requires--e.g., remains undistracted by irrelevant activities.

(9) Utilizes available free time constructively--e.g., initiates productive activity when there is no ongoing class activity.

(11) Withstands interference while engaged in individual activities, as in doing difficult assignment, puzzle, painting, etc.--e.g., does not abandon task when interrupted.

(14) Takes appropriate initiative in problem-solving situations--e.g., asks questions for information which seem to be prompted by general curiosity rather than bids for attention; seeks needed help; asks questions in order to obtain further information or clarification for task at hand.

(20) Tries to figure out things for himself before calling on teacher (leader) or other children for help--e.g., tries to overcome obstacles in the environment.

(21) Enters readily into new situations; is venturesome, inquisitive, etc., when appropriate to ongoing activities--e.g., is eager to begin new and different activities.

(25) Attempts to resolve difficulties that arise between himself and other children without appealing to teacher (leader)--e.g., does not initially seek adult intervention to solve difficulties between himself and peers.

Dependence Items--9

(1) Withdraws in the face of difficulty--e.g., abandons goal when presented with an obstacle.

(12) Imitates others rather than initiating own activities--e.g., changes own behavior to conform to the influence of other children.

(13) Gets upset when called upon in class; appears fearful, tense, timid--e.g., requires repeated urging before making a verbal response.

(15) Seeks to be near others--e.g., follows teacher, assistant, leader, classmates about the room.

(17) Requires close assistance or direction in order to work at a task--e.g., easily distracted; must have teacher or assistant's help to remain attentive to task.

(19) Lets other children impose on him or boss him around--e.g., does not stand up for his rights; yields his place in line, relinquishes turn at play, etc.; allows others to exploit him.

(23) Seeks constant recognition and/or approval--e.g., tries to get teacher's or assistant's attention as frequently as can.

(26) Approaches new tasks timidly and without confidence, shrinks from trying new things--e.g., requires strong encouragement before participating in activities.

(27) Tends to choose habitual and familiar activities.

Pseudoindependence Items--9

(2) Is inattentive, is easily distracted by things going on around him--e.g., engages in random class-unrelated activity (wanders about, daydreams when should be working on own); talks to visitors about irrelevant matters when he/she should be engaged in a task; daydreams; attention wanders from task at hand; is not prepared to answer when called upon in class.

(3) Uses materials and equipment in a rough or destructive manner--e.g., destroys books, crayons, games etc.

(7) Brags, shows off, displays exaggerated opinion of own abilities.

(8) Seeks attention, i.e., in a manner that seems designed for active notice rather than stemming from a need for help--e.g., is inappropriately aggressive in seeking the attention of adults.

(10) Interferes with the work of others--e.g., imposes on other children and bosses them around; inappropriately insists on being admitted to a small peer group; destroys or takes away the work or materials of other children.

(16) Does not follow directions--e.g., works or plays by self despite teacher's request that all children engage in group activity.

(18) Seeks to do things differently from others, even when own method is not effective--e.g., persists at tasks which are clearly impossible.

(22) Does not pursue task to completion--e.g., leaves task at hand

for another or to wander about the room.

(24) Speaks out of turn; interrupts others; takes others' turn, etc. --e.g., refuses to wait his turn.

The Behavioral Transfer Sessions

We had originally planned to eliminate from the second phase of the study any children for whom unambiguous ratings of Independent, Dependent, or Pseudoindependent could not be made. Since the two independent raters had not completed their ratings of the Ss (on the basis of data collected in the classroom observations) and the Z-score checklist designations (see next chapter) also were not completed by the time scheduled for running the transfer sessions, the entire sample was included in the second (transfer) phase of the study. In the final analysis of data, however, all children for whom it was not possible to make an unambiguous rating were excluded.

Transfer session leaders and observers were trained during pilot-testing of the tasks on second-grade children at Public Schools 79 and 175. The leaders were two experienced psychological testers (one male and one female). Each leader conducted approximately one-half of the experimental sessions at each of the two schools from which the control and experimental samples were drawn (Public Schools 90 and 175).

Four experienced observers were also trained, two to serve regularly, and two to substitute on occasions when it was necessary to schedule sessions simultaneously at both schools. In the latter case, one of the regular observers was always paired with one of the substitute observers.

The two session leaders and the four observers were staff members who were not involved in the collection of data in the first phase of the study; each, of course, did not know the composition of the groups, that is, the experimental or control status of the subjects.

The leaders and the observers were rotated between the schools, and the children were randomly selected for each session at each school.

As noted, there were 39 experimental Ss. Since 39 Ss were also run as controls, the total number of children in the experimental sessions was 78. There were, in all, 15 sessions: seven sessions at Public School 90 (involving 36 children); and 8 sessions at Public School 175 (involving 42 children). Table 6 presents the distribution of experimental and control Ss by sex and school for the transfer sessions.[3]

Parental permission was obtained for working with the control children in this study. The principals of the schools were extremely helpful in guiding us in this procedure to insure the greatest degree of cooperation from parents and teachers of the non-Institute children.

Subjects were run in groups of four or six, the composition of each group being consistently one-half experimental and one-half control children. Other than this criterion, the children were completely randomized, so that the mixture of independent, pseudoindependent, and dependent Institute children varied from group to group. This procedure was adopted to eliminate the possible experimenter bias that might emerge if the observers knew that the composition of the group was always partially structured by the distribution of the experimental Ss.

The original plan had been to match experimental and control Ss by sex. This, however, proved not to be feasible because of difficulties encountered in obtaining parental permission for use of control Ss. Ultimately, to avoid causing excessive hardship to school personnel, it was necessary to drop the requirement for matching control to experimental Ss by sex.

Development of the Transfer Tasks

We devoted considerable staff discussion during the early part of the year to the exploration of various possibilities for tasks, games,

---

[3]See Note, Table 5.

puzzles, and tests to be used in the transfer situation which would allow a variation of behavior to emerge. As a matter of fact, members of our research team made several trips to various commercial dealers of toys and playthings in New York City to obtain ideas for techniques, puzzles, games, etc. which might be appropriate for our needs. Specifically, Creative Playthings and Childcraft were visited as well as several department stores.

The selection of such tasks was by no means a simple procedure. The tasks had to be difficult enough, with children of unequal intellectual capacities, to allow choices for each child in terms of seeking help, asking appropriate questions, working diligently on his own, seeking help even though the solution is reachable, etc. Tasks that were too easy would not allow dependent behavior to emerge. Tasks that were too difficult would cause all children to seek help or to give up. Further, whether the Ss would be observed working together in small groups or alone as they solved the tasks had not yet been resolved.

By the end of October, we had considered, but had neither accepted nor rejected, various tasks and situations including, among many others, the following:

(1) An Asch-type task (Asch, 1956) which would allow conformity-dependency behavior to emerge. In this situation, Ss would have the opportunity of producing written or verbal responses, so that objective measures of conformity-dependency would be available.

(2) Various standardized or semistandardized tests and clinical instruments such as the Block Design test, the Goldstein-Sheerer tests, the Vigotsky test, the Stencil Design test of the Arthur Point Scale, etc.

(3) Various experimental techniques, such as the level of aspiration procedure so popular in the 'forties, in which a child could be induced to conform to, or could withstand, influence to alter his level of aspiration

in accordance with the examiner's instructions (or quasi-information about the subject's performance or the performance of peer groups).

(4) A large jigsaw puzzle of a map of a moderately unfamiliar area, possibly Canada. With this task, the Ss could be instructed to put the pieces together as best they could, on their own. They would be informed, however, that there was a large map of Canada behind a screen which they could consult, if necessary. Persistence and the lack of consultation would be the criteria for best (independent) performance.

The extensive pilot-testing of the tasks to be used in the transfer sessions was based on many of the same second-grade children who were observed during the preliminary work with the behavior checklist, thereby giving us the advantage of knowing their characteristic classroom behavior.

The tasks for the transfer sessions that were finally developed to elicit behavior relevant to the three categories in which we were interested are described in the next section. One of the major criteria for their selection was the range or variety of behaviors they seemed to elicit, including interaction with peers as well as authority figures. In the pilot phases, we also evaluated various tasks from the point of view of ease of task-administration. Finally, during this preliminary period, we assessed the potential of various tasks to elicit rater-agreement as to the Ss' behavior.

We devised some unstructured as well as some structured situations to permit different types of behavior to emerge. The structured situations were generally based on the Asch model (1956), modified to avoid the need for a confederate or "stooge" role. This was necessary since it was not feasible to use the children as confederates. The use of an adult confederate would have confounded the situation; Crandall (1967), for example, reported that earlier research had indicated that compliance to adult

pressures and values was found among high-achieving children of this age group.

The behavioral tasks, which we devised for peer groups of four or six Ss, are described below, as noted. It will be seen that tasks (1), (3), and (4) are relatively structured, while tasks (2) and (5) are relatively unstructured. The latter tasks provide opportunities for moderately wide variations in behavior, while the former do so to a lesser extent.

In the actual administration of the experimental sessions, tasks were presented in rotation so that the order in which they appeared was balanced across all groups. Each session lasted for approximately three-quarters of an hour.

The behavioral sessions were conducted in vacant classrooms in Public Schools 90 and 175. Desks and chairs were arranged in the front of the room so that the subjects could be seated with empty desks between each of them. The raters sat at either end of a large desk placed at the side of the room. Here they could easily observe the children without being in their direct visual field. The materials used for the various tasks were also kept at this desk. Another small desk was situated in front of the room, directly under the blackboard. This desk and the chalk tray were used for displaying materials to the children.

Prior arrangements had been made with the teachers for taking the children from their classrooms. While the leader went to pick up the children, the raters prepared the room and task materials. This procedure insured that the raters remained unaware of which children were experimental Ss and which were control Ss.

During the sessions, the group leader stood at the front of the room, attempting to create as informal an atmosphere as was possible while still maintaining the subjects' attention. As the sessions progressed, the two observer-raters recorded frequencies of various behaviors as well as

comments on worksheets which had been developed and polished during the pilot phases. These observations, special comments, responses, and tallies were used by the observers in their ratings of each S. Immediately after each session, then, the observers independently rated each S on the following scale:

Rater_____ IDS Code No._____

Check only one:

I--Independent	Highly_____	Moderately_____
P--Pseudoindependent	Highly_____	Moderately_____
D--Dependent	Highly_____	Moderately_____
X--Can't rate		

Degree of confidence in rating--check one:

3___high 2___medium 1___low

It should be noted that all tasks were employed in qualitative fashion by the raters who, in effect, were rating the Ss globally.

Administration and Description of the Tasks

After all the subjects were in the room, E said:

> GOOD MORNING (AFTERNOON). I'M MR. (MISS)_____.
> NOW I'D LIKE EACH OF YOU TO INTRODUCE YOURSELF.

E called on each S to give his name and recorded each name for identification purposes. E then said:

> YOU'RE GOING TO BE WORKING ON DIFFERENT KINDS OF GAMES AND THINGS. WHILE I'M GETTING THE THINGS READY, YOU MAY COME AND GET A BOOK TO READ, IF YOU WISH. YOU MAY TAKE IT BACK TO YOUR SEAT.

E pointed to a large stack of books. This introduction helped to ease the Ss into the situation and also occupied them while E was preparing

the materials.

(1) Peer influence--bean guessing.

Jars that were approximately 1½" in diameter, ranging in height from about 2" to 7" and filled with beans were presented in ascending order of size. There were six such jars available, but either four or six were employed depending on the size of the group (i.e., either four or six Ss).

> ALL RIGHT. WE WILL PUT THE BOOKS ASIDE NOW, AND TRY SOME GUESSING GAMES. HERE IS A JAR FILLED WITH BEANS. I'M GOING TO CALL ON EACH OF YOU, IN TURN, TO GUESS HOW MANY BEANS THERE ARE IN THE JAR. NOW, DON'T ANSWER UNTIL YOU'RE CALLED ON. _____ (CHILD #1): HOW MANY BEANS DO YOU THINK THERE ARE IN THIS JAR?

E then called on each child in order.

> VERY GOOD. NOW, HERE IS ANOTHER JAR FILLED WITH BEANS. LET'S SEE HOW WELL YOU CAN GUESS ON THIS ONE. REMEMBER, DON'T TALK UNTIL I CALL ON YOU. THIS TIME _____ (CHILD#2) WILL START. _____, HOW MANY BEANS DO YOU THINK THERE ARE IN THIS JAR?

This procedure was repeated with the jars of different sizes until each of the children had responded in all order positions.

This behavioral situation, an Asch-type structured task, indicated the extent to which S's guesses were influenced by those preceding him; or, conversely, the extent to which his judgments remained independent of guesses made by other children in the group.

(2) Spontaneous behavior, self-reliance, peer-influence--geometric puzzles.[4]

For this task, a seven-piece, colored plastic puzzle was employed. A great variety of geometric designs can be constructed with this puzzle by arranging the pieces in various ways. Although the puzzles appear simple, they are in fact quite difficult to solve. The two designs selected for our task are shown in outline form in Figures 1 and 3, while

---

[4]This is a Kohner Educational Puzzle, called VOODOO, manufactured by Kohner Bros., Inc., P.O. 294, East Paterson, N.J.

their corresponding solutions appear in Figures 2 and 4. All figures are scaled to actual size. The pieces of each puzzle distributed to the Ss were numbered with a marker to correspond to the numbers on the solution diagrams. E introduced this task by saying:

> OKAY, NOW WE'RE GOING TO DO SOMETHING ELSE. I'M GOING TO GIVE EACH OF YOU A PUZZLE, AND I WANT YOU TO MAKE A DESIGN. TAKE THE PIECES OUT OF THE ENVELOPE, BUT DON'T START UNTIL I EXPLAIN EVERYTHING.

While talking, E distributed the envelopes, placed the outline of the first figure on a stand in front of the Ss, and the solution face down on the desk. Pointing to the displayed outline, E said:

> NOW, TO MAKE THIS FIGURE CORRECTLY, YOU MUST USE ALL SEVEN PIECES. TRY TO WORK THE PUZZLE AS FAST AS YOU CAN.

E briefly flashed the solution:

> I'M GOING TO PUT THE PICTURE THAT SHOWS HOW TO WORK THE PUZZLE HERE. I WANT YOU TO TRY YOUR BEST TO DO THE PUZZLE YOURSELF--BUT, IF YOU REALLY THINK YOU CAN'T DO IT, YOU MAY COME UP AND LOOK AT THE PICTURE. OKAY, START NOW, AND WORK AS FAST AS YOU CAN. TRY TO DO THE VERY BEST YOU CAN.

E withdrew to the back of the room for several minutes, and then came forward to ask:

> WOULD ANYONE LIKE SOME HELP WITH THE PUZZLE?

To maintain the children's interest and give them a sense of closure, E made sure that everyone had succeeded in solving puzzle (A) before the second design (B) was introduced. E then said:

> NOW SCRAMBLE ALL THE PIECES OF THE PUZZLE AND WE WILL TRY TO MAKE ANOTHER ONE.

The same procedure was followed for the second design (B).

This task provided such information as: persistence before looking at the solution; whether the child copied from his neighbor in attempting to solve the puzzle; whether the child really worked at the solution or simply did not try; whether the child imitated another child by going up to look at the solution; and whether the child resisted or readily accepted the offer

for adult assistance.

(3) Peer influence--circle, square, star, dot guessing.

Six sets of six cards each were prepared:

Set (A) was constructed of black posterboards 11" x 14". Each card contained approximately 150 stick-on circles of various bright colors. For easy identification, the cards were labeled with the letters F, H, J, L, N, and R. Figure 5 is a partial reproduction of a card in this series.

Set (B) consisted of 11" x 14" blue posterboards each containing stick-on circles of a single color: either green, orange, pink, red, tan, or yellow. Figure 6 is a partial reproduction of the green card.

Set (C) also consisted of 11" x 14" posterboards, three with black and three with grey backgrounds. Each card was pasted with squares of a single color, either blue, green, orange, purple, red, or yellow. Figure 7 shows part of a card from this series.

Set (D) was constructed of gummed stars of various colors on 9" x 11" posterboards. The backgrounds of these cards were red, green, or blue and there were two cards of each. These cards were labeled G, L, O, S, X, and Y. See Figure 8 for an example of this set.

Set (E) contained 8-inch square drawings which were used in another (unrelated) research investigation with children.[5] These consisted of a heavily dotted background containing a more finely dotted figure of an object: clock, cup, gun, hand, key, and safety pin. The cup stimulus is shown in Figure 9.

Set (F), also developed previously, was similar to Set (E) except that the background consisted of a grid instead of dots. The common objects pictured in this series were: boat, horse, pail, pan, rabbit, and scissors (see Figure 10, the horse stimulus).

---

[5] C. Deutsch and F. Schumer. Brain-damaged children: a modality-oriented exploration of performance. New York: Brunner/Mazel, Inc., in press.

Arranging Set (A) for display, E said:

> (A) ALL RIGHT. NOW WE'LL TRY ANOTHER GUESSING GAME. I'M GOING TO PUT THESE CARDS HERE, AND I WANT EACH OF YOU TO GUESS WHICH ONE HAS THE MOST CIRCLES. DON'T ANSWER UNTIL I CALL ON YOU. WHAT I WANT YOU TO DO IS TO GUESS WHICH CARD HAS THE MOST CIRCLES--CARD F, H, J, L, N, OR R.

E called on each child in a predetermined order. He then introduced the five (or three) additional sets of stimuli, calling on the children in a predetermined order so that each child responded in all order positions. (All six sets were used when a full group of six children was present. When the group contained four Ss, only the first four sets of cards were introduced.)

> (B) OKAY, NOW WE'LL TRY ANOTHER GUESSING GAME, WITH SOME DIFFERENT CARDS. EACH OF THESE CARDS HAS CIRCLES OF A DIFFERENT COLOR. BE SURE NOT TO ANSWER UNTIL I CALL ON YOU. WHICH ONE HAS THE MOST CIRCLES: GREEN, ORANGE, PINK, RED, TAN, OR YELLOW? THIS TIME _____ WILL START.

All other sets were introduced with instructions similar to those for (B), except that the Ss had to guess about squares, or stars, or dots.

This is also an Asch-type task, giving indications of conformity to peer influence.

(4) Self-reliance--counting animals.

For this task, 8½" x 11" booklets with drawings of animals and other common objects were constructed. Each page of the booklet contained nine pictures. The total number of animals pictured in each of the six booklets ranged from 48 to 53. Thus, each S would arrive at a different number if he counted the animals correctly. Figure 11 is a page from one of the booklets. E said:

> NOW WE'LL DO SOMETHING ELSE. AFTER I FINISH EXPLAINING YOU CAN START, BUT DON'T START UNTIL I TELL YOU WHAT I'D LIKE YOU TO DO.

E distributed a booklet to each child, face down. E continued:

> I WANT YOU TO GO THROUGH THE BOOKLET AND COUNT ALL OF THE PICTURES OF ANIMALS THAT YOU SEE THERE. IF YOU FINISH COUNTING BEFORE EVERYONE FINISHES, JUST LOOK UP, BUT DON'T SAY ANYTHING. SIT QUIETLY AND WAIT UNTIL EVERYONE FINISHES, SO THAT YOU WON'T INTERFERE WITH THEIR COUNTING. AFTER YOU HAVE ALL COUNTED THE ANIMALS, I'LL ASK YOU TO TELL ME HOW MANY YOU COUNTED.

After all the children had finished counting, E called on them, one at a time, for their estimates. E then continued:

> NOW I WOULD LIKE TO BE SURE THAT YOU COUNTED CORRECTLY, SO IF YOU DON'T TRUST YOUR OWN COUNTING, YOU MAY ASK SOMEONE ELSE TO CHECK IT OVER FOR YOU.

This task is another variation of the Asch-type situation, intended to elicit measures of self-reliance.

(5) Authority and/or peer influence--picture construction.

The materials for this task consisted of sheets of colored geometric "stick-ons" and 5" x 8" index cards. A simple doll-like figure using a few of the stick-on shapes given to the children had been constructed for demonstration purposes (see Figure 12).

Displaying the preconstructed design, E said:

> NOW I'M GOING TO GIVE EACH OF YOU A CARD AND SOME OF THESE COLORED SHAPES, AND YOU CAN EACH MAKE A PICTURE. JUST TO SHOW YOU HOW THESE SHAPES CAN BE USED, HERE IS A PICTURE SOMEONE ELSE DID. YOU CAN MAKE ANY PICTURE YOU LIKE. WHEN YOU'RE FINISHED, TURN THE CARD OVER AND PUT YOUR NAME ON IT. I WANT TO KEEP THE CARDS FOR AWHILE, BUT THEN I'LL GIVE THEM TO YOUR TEACHER SO THAT SHE CAN RETURN THEM TO YOU.

This task was designed to indicate S's conformity to authority and peer influence. The Ss could either duplicate the E's picture, copy from their neighbors, or construct their own pictures. The pictures were returned to the Ss after the year's work was completed to avoid contamination of responses to this task.

The next chapter presents the findings based on the research methods described in the current chapter.

Chapter 3

Findings

The material in the present chapter is presented, roughly speaking, in two sections--reliability findings and the development of certain scores; and results of data analyses pertaining to the testing of hypotheses.[6]

I

Reliability Measures and the Development of Scores

This section consists of the following: (a) checklist--reliability of observations for paired observers for each item; (b) checklist--item-total correlations (within each category); (c) checklist--the development of quantitative, continuous scores yielding subject classification procedures (based on Z-scores); (d) reliability of "global" ratings based on checklist--two independent raters; (e) reliability of ratings in the transfer sessions--two independent raters; and (f) checklist--the relationship of Z-score designations and "global" ratings.

Checklist--Reliability of Observations for Paired Observers for Each Item

As will be recalled, and as can be seen in Appendix A, the Classroom Behavior Checklist comprised items which fell into the three categories of independence, pseudoindependence, or dependence. Observations of these behaviors for each day of observation for each child (over a period of time) were tallied for the different frequency points: frequently, moderately often, rarely, and never.

To calculate the reliability of observations per item, for each S a frequency count was made of the number of tallies per item for each of the foregoing frequency points for each of the observation sessions, with

[6] See Edwards (1954) for a description of the various statistical procedures we employed.

data for each observer treated separately. The can't rate point was excluded in this procedure. Totals of these counts for each child per item across all of the observation sessions were then calculated. These computations yielded the total number of "ratable" time-sampling observations made for each child for all sessions for each observer. To obtain a score for each child, weights were assigned: frequently--3; moderately often--2; rarely--1; and never--0. (Higher scores for any item reflect "more" of the particular behavior observed, i.e., greater independence, greater pseudoindependence, or greater dependence.) These weights were then multiplied by the number of observations for the scale point, yielding a total score for that scale point. These scores were then totaled across scale points. Totals were then divided by the number of ratable observations to yield a mean score for the child, thereby eliminating irrelevant variance resulting from unequal numbers of observations. These item total scores were finally transformed by adding a constant (+1) to each, to eliminate zero scores and products in determining inter-observer reliability.

The total scores for each item were paired for each independent observer, and Pearson correlation coefficients were then computed for these paired scores. The coefficients thus obtained indicate the amount of agreement between observers. These correlation coefficients are reported in Table 7.

An examination of Table 7 discloses a wide range in magnitude of the obtained coefficients--with the pseudoindependence category containing the largest number of items with higher reliabilities. Because reports by the classroom observers indicated extreme difficulty in working with some of the items (some relevant behaviors simply did not occur), we decided to systematically explore the possibility that the items with lower reliability coefficients were those which were simply not ratable or for which there

were no opportunities to observe relevant behavior, while the items with higher coefficients were those which related to behaviors which occurred with greater frequencies. Accordingly, we arbitrarily established coefficient cut-offs of .70 and above and .30 and below, and tallied frequencies of the no opportunity to observe category for each item thus designated as possessing high or low reliability. The results, presented in Table 8, consistently indicate that our expectations were correct. Coefficients of .70 and above were indeed found for items that were associated with consistently observable behavior, while coefficients of .30 and below were found for items that were associated with behavior that simply did not emerge and therefore could not be rated. Interestingly, items with coefficients of .30 and below (n=5) were subsumed in the independence and dependence categories, but not in the pseudoindependence category--the latter apparently reflecting behaviors that are salient, observable, ratable, and moderately reliable in terms of observer agreement.

The next step was to eliminate from the pool of checklist items those with low inter-rater reliability. We chose a cut-off of coefficients of .30 or less for this procedure. Accordingly, all items not reaching the criterion of a greater than .30 coefficient of reliability were eliminated from the checklist. By this criterion, eliminated at this stage were the following five checklist items (note--no checklist item from the pseudoindependence category had to be eliminated on the basis of this criterion):

Independence Item 14: takes appropriate initiative in problem-solving situations (r=.18)

Item 21: enters readily into new situations; is venturesome, inquisitive, etc., when appropriate to ongoing activities (r=.02)

Dependence Item 12: imitates others rather than initiating own activities (r=.12)

Item 13: gets upset when called upon in class; appears fearful, tense, timid (r=.11)

Item 26: approaches new tasks timidly and without confidence; shrinks from trying new things (r=.15)

Checklist--Item-Total Correlations (Within Each Category)

Scores for retained items were next correlated with total category scores, using the Pearson correlation method. (Nunnally [1967] has recommended that if multipoint items are employed, the regular Pearson product-moment correlation method is appropriate.)

Table 9 presents the resulting series of item-total correlation coefficients between a particular item and the total for the particular category in which it is subsumed. From this table, it can be seen that several of the items (none in the pseudoindependence category) fail to reach the criterion of significance of p is .05 or less for the particular coefficient. These items are:

Independence Item 4: volunteers contributions to class discussions and projects

Dependence Item 15: seeks to be near others

Item 27: tends to choose habitual and familiar activities

The remaining items achieve high levels of item-total correlation. These items comprise a final pool which constitutes the new, revised checklist. This new checklist, then, presented as Appendix B, contains only

those items which met the criteria of sufficiently high inter-rater reliability and high item-total correlation coefficients to warrant their continued inclusion in the final pool. The revised checklist (consisting now of six items in the independence scale, four items in the dependence scale, and nine /the original group/ items in the pseudoindependence scale) is being used in a continuation of the current study to explore correlates and construct validity of the behaviors which we are able to measure reliably. This includes explorations of such variables as creativity, sociometric status, and self-concept.

Quantitative, Continuous Scores for Checklist--Z-Score Classification

It became apparent as we were working with the checklist material that it would be desirable to devise scoring procedures wherein quantitative continuous scores (within each category) could be calculated for each observed S. This would permit us, for current as well as future purposes, to be more flexible with regard to using various data-analysis procedures, for example, in ranking subjects within each of the categories, or in working with extreme subjects (determined quantitatively, rather than nominally) within a particular category.

After considering several possibilities, the procedure selected for the "scoring system" for the checklist was based on computing Z-scores (separately for each school) for each S for each item within the checklist categories (independence, dependence, pseudoindependence) based on the scoring system outlined in detail in an earlier section of this chapter. Note, Z-scores were based only on items retained after the foregoing reliability and item-analysis procedures were completed. Further, since two scores for each item per S were available (there were two observers), the procedure for obtaining a single Z-score involved computing the mean of the two observer's

item score as described in the section noted above.

Within each school Z-scores were calculated per item for each S. Mean Z-scores were next computed for items within each category for each S. Table 10 presents the three mean Z-scores for each S, along with his rank in each category.[7] Table 11 presents the Ss' ranks in all three categories with Ss ordered according to rank on the independence mean Z-scores from high to low.

A "rule-of-thumb" was then devised for designating subjects for purposes of data analysis on the basis of their ranks on the three mean Z-scores as presented in Table 11. Subjects with ranks of 1 through 20 on the independence items (above median rank) are to be regarded as "independents," provided that their ranks on both of the other categories fall below their respective medians, that is, fall between 21 and 40. Subjects with ranks in the independence category of 21 through 40 are to be regarded as "nonindependents," provided that their ranks for one or the other or both of the two remaining categories, dependence or pseudoindependence fall at or above the median of 20 (i.e., 1-20). On the basis of these criteria, the designations of the Ss as presented in Table 11 resulted in classifying 13 Ss as "independent," 18 as "nonindependent," and 9 as "mixed" types.

---

[7] The data in Table 10 are presented separately according to school. Were there only chance differences in ranks due to schools, then mean of ranks for designation within each school would be 20 or as close to 20 as is possible. In actual fact, these means do indicate this to be the case. That is, as expected, there seem to be little or no school differences in assigned rank within each designation. Specifically, these means are:

School	Designation	Mean of Ranks
P.S. 90	Independence	20.50
	Dependence	20.59
(N=22)	Pseudoindependence	20.86
P.S. 175	Independence	20.50
	Dependence	20.33
(N=18)	Pseudoindependence	20.06

Reliability of "Global" Ratings Based on Checklist--Two Independent Raters

Simultaneously with, and independently of, the statistical determination of observer and item reliabilities, overall ratings for each child were made by two additional, independent raters (the Principal and Co-Principal Investigators), using the scale presented below. These raters based their ratings on two summary checklist sheets for each child (one for each independent classroom observer) which contained total frequencies for each scale point for each item on the behavior checklist. As noted, it was not possible to use the "narratives" (comments by the observers on the child's classroom behavior with regard to anything salient that might have occurred) in this process, since they did not contain sufficiently complete or consistent information from child to child to warrant their use.

Rater_____ IDS Code No._____

Check one only:

I--Independent	Highly_____	Moderately_____
P--Pseudoindependent	Highly_____	Moderately_____
D--Dependent	Highly_____	Moderately_____

X--Can't rate_____

Degree of confidence in rating--check one:

3_____high 2_____medium 1_____low

After these ratings were completed, inter-rater reliability was determined on the basis of Cohen's recommended use of weighted kappa (1968) as a coefficient of agreement for nominal scales. This measure not only

corrects for chance agreement, but also allows for differential weighting of disagreements according to the degree of gravity of such disagreements. It is therefore possible to weight scale-point disagreements (i.e., highly, moderately) less heavily than categorical disagreements (i.e., independent, pseudoindependent, dependent). Table 12 contains the matrix of the weights we employed for this procedure. These weights range from 0 to 4. The larger the weight, the greater the disagreement; 0 indicates no disagreement, and 4 indicates maximum disagreement.

Table 13 shows the distribution of the global ratings made by the independent raters. These data yielded a reliability coefficient of .76 when weighted kappa was computed, taking into account scale points as well as category. This results in $z=5.43$ ($p < .0002$, two-tailed). (Unweighted kappa, based on categorical designation only, yielded a reliability coefficient of .873, resulting in $z=4.48$ [$p < .0002$, two-tailed].) It can therefore be concluded that the raters showed excellent agreement in designating the Ss into the three relevant classifications on the basis of the summary checklist data.

From Table 13, it can be seen that global ratings based on checklist summary information tend to yield little disagreement. It should also be noted that both raters were able to execute the ratings quickly and with considerable confidence. By and large, even ratings made with low confidence tended to agree with each other. The raters were under the general impression that the degree of confidence each rater tended to assign to her ratings was a question of personal "style"--for there seemed to be differences between the raters in the frequency with which "low confidence" was checked.

Reliability of Ratings--Transfer Sessions

Chapter 2 describes the manner in which ratings of behavior in the transfer sessions were made. Inter-rater agreement (for the scale just presented) was indexed by means of Cohen's weighted kappa (1968). As already noted, this computation not only corrects for chance agreement, but also allows for differential weighting of disagreements according to the degree of gravity of such disagreements (see Table 12).

Table 14 shows the frequency distribution of the ratings by paired observers of the transfer sessions. These data yielded a reliability coefficient (kappa) of .257, which, in terms of the normal curve deviate, results in $z=2.18$ ($p < .03$, two-tailed). This coefficient is based on the 46 cases remaining after the elimination of ratings in which observers expressed low confidence ($N=24$) and those designated "can't rate" ($N=6$).[9]

Table 15 summarizes the number of agreements and disagreements within and across categories. As can be seen, most of the disagreements were in regard to the independence--dependence ratings, and among these, most were in regard to the moderately independent vs. moderately dependent choices.

It is thus seen that observers' agreement in the transfer sessions, although moderately good, does not fall on the same level as that obtained for a different set of raters who made judgments on the basis of the checklist. It is highly probable that the kinds of operations involved in making such ratings differ considerably in each situation, with judgments in the transfer situations more liable to various types of error than those based

[9] Inclusion of "low confidence" ratings in the computation of weighted kappa resulted in a rather low reliability estimate prompting us to exclude those cases from the analysis described. When these cases are included (total $N=70$, for there were 6 "can't rate" designations), the findings were: weighted kappa was .182, which in terms of the normal curve deviate, results in $z=1.89$ ($p < .06$, two-tailed).

on checklist information.

Checklist--The Relationship of Z-score Designations and "Global" Ratings

When ratings based on qualitative examination of the checklist were completed, they were compared to the designations of the experimental Ss based on their Z-score ranks for each of the categories on the checklist. Note, for this analysis, those Z-score classifications which were labeled "mixed" and those global ratings which represented disagreements between the raters or which involved a "can't rate" designation were notused. This analysis then, was based on a group of 29 Ss.

Table 16 presents the contingency classification of designations derived from each of the above sources. A chi-square analysis as to whether these two methods of classifying the Ss are associated yielded a probability level of $<.025$ and a contingency coefficient of .486. It is thus concluded that there is a positive association, established with some degree of confidence, between the method of rating behavior globally (on the basis of qualitative examination of the checklist material) and that of objectively classifying behavior on the basis of the Ss' Z-scores.

II

Data Analyses Pertaining to the Testing of Hypotheses

The current section describes findings pertaining to the following analyses: (a) global checklist ratings vs. ratings (experimental subjects only) in the transfer situation; (b) Z-score checklist designations vs. ratings in the transfer situation; and (c) distribution of experimental Ss according to "filler" status vs. global checklist ratings, Z-score checklist designations, and behavioral (transfer session) ratings. The foregoing is Part A of this section, and relates to explorations involving the experimental

Ss only. Part B of the current section pertains to experimental vs. control differences in terms of the ratings of independence or dependence made in the behavioral sessions. Analyses of results were made for schools separately and combined as well as for groups with and without Ss for whom ratings with "low confidence" were made.

A. Experimental Ss Only

Global Checklist Ratings vs. Ratings in the Transfer Situation

The next two sections present data analyses for only those Ss (experimental Ss) observed in the classrooms. These sections are addressed to the question regarding "transfer"--that is, whether designations based on classroom behavior are in any way associated with designations based on behavior in the nonclassroom situation. The first of these analyses is concerned with whether the global ratings based on the checklist are associated with status as rated in a later transfer (behavioral) situation. The findings, as presented in Table 17, indicate a lack of association between these two methods of classifying the Ss.

Z-Score Checklist Designations vs. Ratings in the Transfer Situation

Table 18 presents data pertaining to the relationship between objective Z-score classifications based on the checklist and ratings made subsequently in a nonclassroom behavioral situation. These findings also indicate a lack of association between these two methods of classifying the Ss.

The foregoing analyses (see Tables 17 and 18) do indeed suggest that insofar as the current research design in concerned, behavior as classified in the classroom is not similarly classified in subsequent nonclassroom behavioral situations. Thus, it would be reasonable to suggest that such behavior does not "transfer." The situations are so different from each

other, however, that one can raise a reasonable doubt as to whether transfer effects were actually being measured by the behavioral sessions. The next chapter briefly touches on this issue again.

"Filler" Status and Three Different Ratings of Experimental Ss

As will be recalled (see Table 1), the experimental children were classified according to "filler" status, depending on the number of years of membership in the Institute classes. One of the hypotheses of this study pertains to the possibility that length of exposure to our demonstration program would be associated with status (independence, pseudoindependence, or dependence) as determined through checklist ratings, Z-score classifications, and behavioral session ratings. (See Table 19 for these frequency distributions.) Table 20 presents results of analyses pertaining to these expectations. All analyses yielded nonsignificant results, it should be noted.

Specifically, all Ss were allocated to one of five groups defined by the year the child entered the program. For example, if an S entered the Institute's program in 1964 at the prekindergarten level, he received four full years of exposure to an enriched classroom atmosphere and curriculum; if he entered the program in 1968 (an F3 subject), however, he virtually had no exposure to the Institute's program when the current study was initiated. In this sense, then, F3 subjects are similar to control subjects, at least in the beginning of the fall term of 1968. To explore the association between various measures and status of subjects in terms of length of exposure to the Institute's program, we grouped all Ss with "filler" classifications of: the full four years of exposure, through three, two, and one year of exposure (E, FK, F1, and F2 Ss) and examined their ratings in comparison to the F3 group. That is, Ss receiving one year or more of our

program were compared to Ss who had entered the program the year of our investigation in terms of distributions of ratings and Z-score designations. As presented in Tables 19 and 20, data analyses indicate that there is no association between "filler" status and global checklist ratings, Z-score classifications, or behavioral session ratings.

B. Experimental vs. Control Ss (Behavioral Sessions)

Ratings of Ss in the behavioral sessions were examined in terms of experimental vs. control status of the subjects. Table 21 presents the results of chi-square analyses for subjects in both schools combined with (a) "low confidence" ratings included and (b) "low confidence" ratings eliminated. Table 21 also presents results for both schools held separately, with ratings of "low confidence" included with these analyses because of the small Ns involved. All analyses were based on four-fold tables (one degree of freedom) with Yates-corrected values of chi-square, and all comparisons involving the experimental group were based only on those Ss with one or more years of exposure to the Institute program.

For all comparisons, the N of the experimental sample was reduced from 39 to 37, because 2 Ss from Public School 175 were randomly eliminated to equate the Ns of the experimental and control samples. (These Ss, as it turned out, were also rater-disagreement Ss.) In comparisons (1) and (2), 9 F3 Ss and 13 rater-disagreement Ss were eliminated from the experimental group; in addition, 5 "low confidence" rating Ss were eliminated from the experimental group in comparison (2).

There are 37 control Ss in comparisons (1) and (2) of which 18 Ss were eliminated because they represented rater-disagreements. In addition, 4 "low confidence" control Ss were eliminated in comparison (2).

In comparison (3) there were 19 experimental and 19 control Ss controlled

(Public School 175). The 9 Ss who were rater-disagreements and 3 F3 Ss were eliminated from the experimental group, leaving a total of 7 experimental Ss. From the control group, 11 rater-disagreements were eliminated, leaving a total of 8 control Ss.

In comparison (4), there were 18 experimental and 18 control Ss (Public School 90). Of the 18 experimental Ss, 4 rater-disagreements and 6 F3 Ss were eliminated leaving a total of 8 Ss. Of the 18 control Ss, 7 rater-disagreements were eliminated, leaving a total of 11 Ss.

All findings reported in Table 21 are nonsignificant--that is, there is no apparent association between experimental and control subject status and the ratings assigned to them based on behavior in the transfer sessions.

Chapter 4

Summary and Conclusions; Abstract of Continuation Research

This chapter will briefly recapitulate the major findings and relate these to the expectations and hypotheses described in Chapter 1. It might be noted that since the currently described research has continued into the present, and is still ongoing, major conclusions and discussion will be deferred until the present year's work is completed. This year's work (1969-1970) involves cross-validation of the checklist with a similar sample, development of reliable measures of creativity, self-concept, and sociometric choice, and exploration of the relationship of scores from the latter three areas as well as teacher-ratings of independence, to more objective measures of independence derived from the checklist. The end of this chapter presents a brief abstract of the current continuation research. It should be noted that this work is well under way.

A basic, overall expectation was that we could develop a reliable, usable classroom behavior checklist. This expectation has been met. Inter-rater reliability coefficients (see Table 7) were utilized to eliminate items of low reliability; in addition, item-total correlations for checklist items within each category (see Table 9) were further examined to eliminate additional items. The resulting checklist, presented in Appendix B, thus contains only those items which met the criteria of sufficiently high inter-rater reliability and high item-total correlation coefficients to warrant their continued inclusion in the final form of the checklist. The revised checklist is being used in a continuation of the currently described study, as noted, to explore correlates and construct validity of independent behavior.

The preceding chapter described the development of an objective Z-score based on checklist findings as well as a systematic "rule" for classifying subjects based on their Z-score profiles. These designations are presented in Table 11, and resulted in classifying 13 Ss as "independent," 18 as "nonindependent," and 9 as "mixed" types. (The latter, it should be noted, is not synonymous with the "pseudoindependent" category.) Global raters (who were not in the classroom) independently rated the Ss on the basis of summary checklist data (see Table 13) with an extremely high degree of reliability--that is, with little or no disagreement.

Table 16 presents data pertaining to the association between global ratings and Z-score designations of the Ss. It was concluded that there is a positive association, established with some degree of confidence, between these checklist methods of categorizing behavior.

Ratings of observers in the transfer situation (see Table 14) were less reliable, it should be noted, than those based on summary checklist data. Several possibilities can be offered to "explain" the latter finding: (a) four different raters (rather than two, as planned) were used in the transfer session; (b) ratings in the transfer session were based on one observation session, whereas data obtained in the classroom were obtained over many sessions of observation; (c) the tasks and behavioral situations we utilized were in many ways unique to the observers and to the subjects alike; despite training, the raters were perhaps not comfortable in assigning ratings to the Ss based on such an unconventional approach; and (d) the nature of the behavioral session was such that it might simply not have elicited reliably ratable behavior (at least to the same extent as did the classroom situation).

For the experimental Ss, overall findings indicate that the designations based on classroom behavior are not associated with designations derived from the transfer situation. Thus, Tables 17 and 18 indicate that insofar as the current research is concerned, behavior classified in the classroom is not "carried over" to nonclassroom behavioral situations. Based on classroom data alone, however (see Table 13), experimental subjects were judged more frequently by independent raters to be independent than otherwise; further, as this table also shows, such raters used the category of pseudoindependence even more frequently than dependence, suggesting that the pseudoindependent category, as expected, is meaningful in terms of characterizing classroom behavior. Also , classroom observers were particularly reliable in judging pseudoindependent behavior; and item-total Pearson correlation coefficients in the pseudoindependent category were all significant.

As presented in Tables 19 and 20, data analyses indicate that there is no association between "filler" status and the global checklist ratings, Z-score classifications, and behavioral session ratings of the experimental subjects. Finally, Table 21 indicates that there is no apparent association between experimental and control subject-status and the ratings assigned to them based on behavior in the behavioral sessions.

Overall, the foregoing findings indicate lack of support for most of the hypotheses and expectations as presented in Chapter 1, with the exception of those pertaining to the checklist, and its development and use. Indeed, it does seem that the "transfer" session was an inadequate method for measuring "transfer" effects, and that although we are seemingly eliciting independent behavior in the classroom, the experimental Ss, when compared to experimental F3 "filler" Ss and to the control group Ss, did not appear to differ in terms of their assigned ratings with any degree of statistical confidence. Some reasons for greater unreliability of the

ratings in these situations were presented earlier. It may also be that the significant and crucial variable is the striking differences between the classroom situation and the behavioral situation. In some respects, the behavioral situation resembled the typical control classroom more than did the experimental classroom. In this sense, Institute independent children might well have not "carried" their independent behavior over into the new situation. Conversely, control Ss might have felt freer to show independent behavior in a situation which was structured to elicit such behavior more than did their typical classroom situation.

Although a complete discussion of these points will be presented in the 1970 Final Report, it should be noted here that the literature presents some evidence that dependency may not be a unitary trait but rather is a rubric for different behaviors. Thus, a child would be rated differently depending on the aspect of dependency measured. Furthermore, some studies indicate that the situational aspects in which the dependency is measured are of importance. It is likely that this evidence is relevant to the study of independence, too.

In any event, we may be forced to conclude that independence does not "carry over" into other situations--at least in this age group--regardless of that situation. Further, the nature of the task, situation, and the requirement placed before the child, seem to be of key significance. Questions of construct validity, that is, just what independence is and what its correlates are seem to be crucial, at this point of our studies, and it is to these issues that our current work is addressed. We are content that we have developed a usable instrument for doing so, at least in terms of classroom activity. Further, since many "conventional" classrooms (kindergarten, for example, or prekindergarten) provide the kind of milieu

that allows observable behavior to emerge (of the sort our checklist is designed to measure), we are hopeful that our cross-validated checklist can be useful at younger age-levels.

Abstract of Continuation Research (1969-1970)

(1) Objectives. In our original study, we proposed to assess the degree to which the Institute's ongoing enrichment program in several Harlem public schools has succeeded in fostering and encouraging independent, autonomous behavior among its pupils. This evaluation has been conducted through the use of behavioral, rather than standardized test criteria, both in the classroom and in a "transfer" situation, to explore whether or not the changes that occur in the classroom do emerge outside of that "programed" setting. This investigation, and the proposed continuation research, have implications not only for scholastic achievement but also for the child's overall adjustment in daily life. If we are indeed training for, and reinforcing independent behavior in Institute children, we must be assured that independence should be an educational goal. For the above reason, we proposed to continue our current research to determine what associated, correlative behaviors and characteristics accompany independent, pseudo-independent, or dependent behavior. The specific objectives of this continuation research are the following:

(a) To replicate and cross-validate the revised classroom behavior checklist with a new but similar pupil sample. This replication would hopefully enable us to provide the professional community with a reliable, cross-validated instrument for the measurement of independent, pseudoindependent, and dependent behavior.

(b) To explore the relationship of teachers' ratings of children

as to independent, pseudoindependent, and dependent behavior to behavioral checklist scores based on observations of these same children.

(c) To explore the sociometric concomitants of independence, pseudoindependence, and dependence.

(d) To explore the relationship of creativity measures and scores to designations of independence, pseudoindependence, and dependence derived from the checklist, and to develop new methods for assessing the creativity dimensions in this sample.

(e) To explore the relationship of self-concept (as measured by Q-methodology) to designations of independence, pseudoindependence, and dependence derived from the checklist.

(f) To examine a number of additional, subsidiary relationships that can shed light on the behaviors relevant to the area of our investigation.

(2) Procedures. Our sample is composed of all children in as many 1969-1970 third-grade Institute demonstration classrooms as are available. Third-graders have been chosen as these children have had the longest exposure to the Institute program and should, therefore, show the greatest effects of this intervention. Within our sample will be "fillers," those children who have been added to the program because of attrition in the initial group. The presence of "fillers" will enable us to determine whether behavioral change is a function of the length of time the child has spent in the program. Using the revised behavior checklist, we plan to proceed as follows:

(a) Early in the school year, two observers will independently record, by frequency tallies on the checklist, each child's behavior during five-minute periods over a total of 20 sessions on different days (this step is currently near completion).

(b) An initial set of sociometric measures will be obtained in each classroom by personnel other than those making the behavioral observations.

This step has been completed. These same measures will be collected again at the end of the year.

(c) Teachers' ratings of the same children in regard to independence, pseudoindependence, and dependence will be obtained at both the beginning and the end of the school year. The first wave of teachers' ratings has been obtained.

(d) Measures of creativity will be developed and administered to the Ss in the sample.

(e) Q-sort techniques will be developed and administered to all of the Ss in the samples simultaneously with step (d) to measure self-concepts and self-perceptions.

At each step along the way, it might be noted, stress will be placed on developing reliable and objective instruments and measures.

References

Asch, S.E. Studies of independence and conformity: I. a minority of one against a unanimous majority. Psychological Monographs, 1956, 70, No. 9 (Whole No. 416).

Beller, E.K. Dependency and independency in young children. Journal of Genetic Psychology, 1955, 87, 23-35.

Cervenka, E.J. The measurement of bilingualism and bicultural socialization of the child in the school setting: the development of instruments. Final report on Headstart Evaluation and Research 1966-67 (contract #66-1).

Cohen, J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, 1968, 70, 213-220.

Crandall, V.C. Achievement behavior in young children. In W. Hartup and L. Smothergill (Eds.), The young child. Washington: NAEYC, 1967.

Davidson, H.H., & Greenberg, J.W. Traits of school achievers from a deprived background. New York: Simon and Schuster, 1967.

Deutsch, C., and Schumer, F. Brain-damaged children: a modality-oriented exploration of performance. New York: Brunner/Mazel, in press.

Edwards, A. L. Statistical methods for the behavioral sciences. New York: Holt, Rinehart, and Winston, 1954.

Eells, K., Davis, A., Havighurst, R.J., Herrick, V.E., & Tyler, R.W. Intelligence and cultural differences: a study of cultural learning and problem-solving. Chicago: University of Chicago Press, 1951.

Finney, J.C. Some maternal influences on children's personality and character. Genetic Psychology Monographs, 1961, 63, 199-278.

Gewirtz, J.L. A factor analysis of some attention-seeking behaviors in young children. Child Development, 1956, 27, 17-36.

Guidelines for testing minority group children. Journal of Social Issues, 1964, 20, 129-145.

Hartup, W.W. Dependence and independence. In H.W. Stevenson (Ed.), Child psychology: The 62nd yearbook of the National Society for the Study of Education, Part I. Chicago: The National Society for the Study of Education, 1963.

Minuchin, S., Montalvo, B., Guerney, B.G., Jr., Rosman, B L., & Schumer, F. Families of the slums: an exploration of their structure and treatment. New York: Basic Books, Inc., 1967.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1967.

Powledge, F. To change a child: a report on the Institute for Developmental Studies. Chicago: Quadrangle Books, 1967.

Zubin, J., Eron, L.D., & Schumer, F. An experimental approach to projective techniques. New York: John Wiley and Sons, Inc., 1965.

Table 1

Experimental Sample Observed in the Classrooms: Mean Age and Sex for Subjects in Public Schools 90 and 175 Classified According to Years of Exposure to the Enrichment Program

Entered IDS Classes	N	Sex	Classification[a]	Mean Age[b] (Months)
		Public School 90		
Pre-K, 1964-1965 (4 years of exposure)	4	M--4 F--0	E	97.25
Kindergarten, 1965-1966 (3 years of exposure)	2	M--1 F--1	FK	101.00
First Grade, 1966-1967 (2 years of exposure)	6	M--3 F--3	F1	99.50
Second Grade, 1967-1968 (1 Year of exposure)	1	M--0 F--1	F2	103.00
Third Grade, 1968-1969 (0 years of exposure)	6	M--3 F--3	F3	98.17
Total, P.S. 90	19	M--11 F--8		98.84
		Public School 175		
Pre-K, 1964-1965 (4 years of exposure)	14	M--8 F--6	E	98.35
First Grade, 1966-1967 (2 years of exposure)	2	M--1 F--1	F1	99.50
Second Grade, 1967-1968 (1 year of exposure)	2	M--2 F--0	F2	99.00
Third Grade, 1968-1969 (0 years of exposure)	3	M--2 F--1	F3	98.00
Total, P.S. 175	21	M--13 F--8		98.48
Total, Schools Combined	40	M--24 F--16	18 Es 2 FKs 8 F1s 3 F2s 9 F3s	98.68

Note--There were no FK subjects in Public School 175 in the third grade.

[a] F categories contain "filler" Ss, added to the experimental group through the years because of attrition.

[b] As of September, 1968. Mean age for all Ss is just under 8 years, 3 months.

Table 2

Experimental Sample Observed in the Classrooms: Mean Age for Subjects Classified According to Years of Exposure to the Enrichment Program for Both Schools Combined

Classification	N	Mean Age[a] (Months)
E	18	98.11
FK	2	101.00
F1	8	99.50
F2	3	100.33
F3	9	97.89
Total	40	98.68

[a] As of September, 1968.

Table 3

Behavioral Transfer Sessions--Experimental Subjects: Mean Age and Sex for Subjects in Public Schools 90 and 175 Classified According to Years of Exposure to the Enrichment Program

Entered IDS Classes	N	Sex	Subject Classification[a]	Mean Age (Months)[b]
		Public School 90		
Pre-K, 1964-1965 (4 years of exposure)	4	M--4 F--0	E	97.25
Kindergarten, 1965-1966 (3 years of exposure)	2	M--1 F--1	FK	101.00
First Grade, 1966-1967 (2 years of exposure)	6	M--3 F--3	F1	99.50
Second Grade, 1967-1968 (1 year of exposure)	1	M--0 F--1	F2	103.00
Third Grade, 1968-1969 (0 years of exposure)	5	M--3 F--2	F3	98.40
Total, P.S. 90	18	M--11 F--7		99.06
		Public School 175		
Pre-K, 1964-1965 (4 years of exposure)	14	M--8 F--6	E	98.35
First Grade, 1966-1967 (2 years of exposure)	2	M--1 F--1	F1	99.50
Second Grade, 1967-1968 (1 year of exposure)	2	M--2 F--0	F2	99.00
Third Grade, 1968-1969 (0 years of exposure)	3	M--2 F--1	F3	98.00
Total, P.S. 175	21	M--13 F--8		98.48
Total, Schools Combined	39	M--24 F--15	18 Es 2 FKs 8 F1s 3 F2s 8 F3s	98.76

Note--There were no FK subjects in P.S. 175 in the third grade.

[a] F categories contain "filler" Ss, added to the experimental group through the years because of attrition.

[b] As of September, 1968. Mean age for all Ss is just under 8 years, 3 months.

Table 4

Experimental Sample in the Behavioral Transfer Sessions: Mean Age for Subjects Classified According to Years of Exposure to the Enrichment Program for Both Schools Combined

Classification	N	Mean Age[a] (Months)
E	18	98.11
FK	2	101.00
F1	8	99.50
F2	3	100.33
F3	9	98.25
Total	39	98.76

[a] As of September, 1968.

Table 5

Behavioral Transfer Sessions--Control Subjects: Mean Age and Sex for Subjects in Public Schools 90 and 175

	N	Sex	Mean Age (Months)[a]
		Public School 90[b]	
	18	M--7 F--11	99.47[b]
		Public School 175	
	19	M--8 F--11	97.84
Total	37	M--15 F--22	98.61[b]

Note--Although 39 Ss were run in the transfer sessions, only 37 were usable for statistical analysis because of certain errors in sample selection. The Ss eliminated were both from Public School 175. To compensate for this reduction in sample size, when experimental and control differences are explored, two Ss from Public School 175 were randomly eliminated from the experimental sample to maintain the initial equality in sample size for the two treatment groups. In statistical analyses involving only the experimental sample, all 39 experimental Ss were used.

[a] As of September, 1968.

[b] Total n from Public School 90 on which mean age is based is 17 subjects because of incomplete school data at this school for one S. Mean age for control Ss is therefore based on 36 cases, and is just under 8 years, 3 months, making control and experimental subjects comparable in this regard (see Table 1).

Table 6

Distribution of Subjects by School and Sex in the Transfer Sessions[a]

Session		Experimental M	Experimental F	Control M	Control F	N
			Public School 90			
1		2	1	1	2	6
2		1	2	1	2	6
3		1	2	1	2	6
4		3	0	1	2	6
5		1	1	1	1	4
6		2	0	1	1	4
7		1	1	1	1	4
	N	11	7	7	11	36
			Public School 175			
1		3	0	1	2	6
2		1	2	1	2	6
3		3	0	1	2	6
4		2	0	2	0	4
5		1	2	1	2	6
6		1	2	3	0	6
7		1	1	0	2	4
8		1	1	1	1	4
	N	13	8	10	11	42
	Total	24	15	17	22	78

[a] See Note, Table 5.

Table 7

Item	Inter-rater Reliability Coefficients of Behavior Checklist Items	Product-Moment Correlation Coefficients[a]
	Independence	
(4)	volunteers contributions to class discussions and projects	.31
(5)	helps, sympathizes, shows consideration of others; is thoughtful; offers praise, guidance, assurance	.35
(6)	sits quietly and attentively when task requires	.71
(9)	utilizes available free time constructively	.50
(11)	withstands interference while engaged in individual activities, e.g., doing difficult assignment, a puzzle, painting	.40
(14)	takes appropriate initiative in problem-solving situations	.18
(20)	tries to figure out things for himself before calling on teacher (leader) or other children for help	.49
(21)	enters readily into new situations; is venturesome, inquisitive, etc., when appropriate to ongoing activities	.02
(25)	attempts to resolve difficulties that arise between himself and other children without appealing to teacher (leader)	.50
	Dependence	
(1)	withdraws in the face of difficulty	.38
(12)	imitates others rather than initiating own activities	.12
(13)	gets upset when called upon in class; appears fearful, tense, timid	.11
(15)	seeks to be near others	.64
(17)	requires close assistance or direction in order to work at a task	.70
(19)	lets other children impose on him or boss him around	.45
(23)	seeks constant recognition and/or approval	.69
(26)	approaches new tasks timidly and without confidence; shrinks from trying new things	.15
(27)	tends to choose habitual and familiar activities	.68
	Pseudoindependence	
(2)	is inattentive; is easily distracted by things going on around him	.54
(3)	uses materials and equipment in a rough or destructive manner	.58
(7)	brags, shows off, displays exaggerated opinion of own abilities	.58
(8)	seeks attention, i.e., in a manner that seems designed for active notice rather than stemming from a need for help	.77
(10)	interferes with the work of others	.72
(16)	does not follow directions	.60
(18)	seeks to do things differently from others even when own method is not effective	.53
(22)	does not pursue task to completion	.35
(24)	speaks out of turn; interrupts others; takes other's turn, etc.	.95

[a] For all reported coefficients, $N=40$.

Table 8

Frequencies of No Opportunity to Observe Category for Behavior Checklist Items with Reliability Coefficients of .30 and Below and .70 and Above

		Frequency--No Opportunity to Observe
	Coefficients of .30 and Below	
	Independence	
(14)	takes appropriate initiative in problem-solving situations	466
(21)	enters readily into new situations; is venturesome, inquisitive, etc., when appropriate to ongoing activities	633
	Dependence	
(12)	imitates others rather than initiating own activities	618
(13)	gets upset when called upon in class; appears fearful, tense, timid	586
(26)	approaches new tasks timidly and without confidence; shrinks from trying new things	601
	Coefficients of .70 and Above	
	Independence	
(6)	sits quietly and attentively when task requires	33
	Dependence	
(17)	requires close assistance or direction in order to work at a task	83
	Pseudoindependence	
(8)	seeks attention, i.e., in a manner that seems designed for active notice rather than stemming from a need for help	36
(10)	interferes with the work of others	35
(24)	speaks out of turn; interrupts others; takes other's turn, etc.	124

Table 9

Item-Total Pearson Product-Moment Coefficients for Checklist Items[a]

Checklist Item No.	r	Corrected[b] r	p[c]
	Independence		
4	.544	.245	n.s.
5	.679	.530	.005
6	.755	.678	.005
9	.692	.579	.005
11	.640	.538	.005
20	.675	.547	.005
25	.622	.344	.025
	Dependence		
1	.700	.518	.005
15	.367	.106	n.s.
17	.804	.661	.005
19	.536	.407	.01
23	.802	.701	.005
27	.603	.262	n.s.
	Pseudoindependence		
2	.766	.678	.005
3	.691	.659	.005
7	.767	.741	.005
8	.825	.768	.005
10	.816	.735	.005
16	.919	.893	.005
18	.606	.506	.005
22	.722	.575	.005
24	.563	.377	.01

[a]Items for which inter-rater reliability coefficients were .30 or less were previously dropped from the checklist.

[b]Corrected for spuriousness (to eliminate item's own contribution to total score).

[c]Items marked n.s. had p values of .05 or higher.

Table 10

Mean Z-Scores and Ranks for Independence, Dependence, and Pseudoindependence Checklist Categories

N=40

Subject Code # P.S. 90	Independence Mean Z-Score	Independence Rank	Dependence Mean Z-Score	Dependence Rank	Pseudoindependence Mean Z-Score	Pseudoindependence Rank
0000	0.08	18	-0.52	16	-1.03	4
937	0.24	20	-0.28	21	1.18	36
939	1.09	36	-1.04	4	-1.31	2
943	0.67	30	-0.56	15	-0.81	7
945	-1.66	3	-0.13	22	0.31	29
947	-1.55	5	2.36	39	0.89	32
948	-1.08	8	0.72	31	0.58	31
955	-0.96	11	-0.11	23	1.39	37
957	0.39	23	0.24	27	0.29	28
958	-0.02	16	-0.72	12	-0.13	25
961	1.03	35	-1.22	2	-1.25	3
962	-1.52	6	0.72	32	0.24	27
967	-0.23	15	-0.72	14	-0.05	26
968	0.75	31	-1.05	3	-0.99	5
1004	-0.96	10	0.86	34	0.33	30
1005	0.60	28	0.20	26	-0.23	21
1010	-0.69	12	-0.52	16	1.13	35
1011	1.24	38	-1.43	1	-1.43	1
1012	1.21	37	-0.74	11	-0.68	10
703	-1.13	7	2.15	40	2.61	39
932	0.34	22	1.12	35	-0.49	17
934	1.45	40	0.66	29	-0.55	14
P.S. 175						
851	1.34	39	-0.49	18	-0.61	11
852	0.08	17	-0.87	8	-0.57	13
854	-1.95	2	1.38	37	2.69	40
855	-0.24	14	0.00	24	1.01	34
872	-1.66	4	0.75	33	2.17	38
875	0.56	26	-0.76	10	-0.55	15
878	1.01	34	-1.00	5	-0.86	6
879	-2.09	1	2.74	38	1.01	33
881	-0.35	13	-0.72	13	-0.18	22
883	-0.97	9	-0.48	19	-0.16	24
884	0.78	32	-0.91	7	-0.69	9
886	0.43	24	-0.36	20	-0.33	20
1003	0.99	33	-0.98	6	-0.59	12
1019	0.32	21	-0.79	9	-0.38	19
1020	0.61	29	0.59	28	-0.42	18
1021	0.58	27	0.05	25	-0.79	8
1023	0.11	19	1.14	36	-0.18	23
1024	0.46	25	0.70	30	-0.55	16

Table 11

Ranks of Checklist Mean Z-Scores for Independence, Dependence, and Pseudoindependence[a]

N=40

Subject Code No.	Rank: Independence	Dependence	Pseudoindependence	Designation[b]
934	1	12	27	M
851	2	23	30	I
1011	3	40	40	I
1012	4	30	31	I
939	5	37	39	I
961	6	39	38	I
878	7	36	35	I
1003	8	35	29	I
884	9	34	32	I
968	10	38	36	I
943	11	26	34	I
1020	12	13	23	M
1005	13	15	20	M
1021	14	16	33	M
875	15	31	26	I
1024	16	11	25	M
886	17	21	21	I
957	18	14	13	M
932	19	6	24	M
1019	20	32	22	I
937	21	20	5	NI
1023	22	4	18	NI
0000	23	24	37	M
852	24	33	28	M
958	25	29	16	NI
967	26	27	15	NI
855	27	17	7	NI
881	28	28	19	NI
1010	29	25	6	NI
955	30	18	4	NI
1004	31	7	11	NI
883	32	22	17	NI
948	33	10	10	NI
703	34	3	2	NI
962	35	9	14	NI
947	36	2	9	NI
872	37	8	3	NI
945	38	19	12	NI
854	39	5	1	NI
879	40	1	8	NI

[a]Subjects are ordered by rank on Independence scale.

[b]Designations are: I = Independent; NI = Nonindependent; and M = Mixed type.

Table 12

Weighting Matrix for Measuring Disagreement

between Paired Observers

	Observer A					
	Independent		Pseudoindependent		Dependent	
erver B	Highly	Moderately	Highly	Moderately	Highly	Moderately
ependent						
ighly	0	1	4	3	4	3
oderately	1	0	3	2	3	2
udoindependent						
ighly	4	3	0	1	4	3
oderately	3	2	1	0	3	2
endent						
ighly	4	3	4	3	0	1
oderately	3	2	3	2	1	0

Table 13

Distribution of Two Independent Raters' "Global" Ratings Based on Summary Checklist Data by Category and Degree

N=40

Category	Agreements	Breakdown of Agreements Within Categories According to Degree		
		High High	High Moderate	Moderate Moderate
Independent	29	13	8	8
Dependent	1	0	0	1
Pseudo-independent	7	0	0	7
	37			
Non-agreements or can't rate [a]	3			
Total N	40			

[a]For two Ss, the raters disagreed on category; for one S there was a can't rate designation.

Table 14

Frequency Distribution of Ratings by Paired Observers of Transfer Sessions[a]

	Observer A						
	Independent		Pseudoindependent		Dependent		
rver B	Highly	Moderately	Highly	Moderately	Highly	Moderately	Total
pendent							
ghly	0	2	0	0	1	1	4
derately	6	10	0	0	1	2	19
doindependent							
ghly	0	0	0	0	0	0	0
derately	0	0	0	0	1	0	1
ndent							
ghly	0	0	0	0	1	3	4
derately	3	5	0	1	1	8	18
tal	9	17	0	1	5	14	46

[a]Includes only those ratings in which the observers expressed high or moderate confidence. Total N was 76 (39 experimental Ss and 37 control Ss). Ratings in which observers expressed low confidence (N=24) were excluded, as are those designated "can't rate" (N=6).

Table 15

Summary of Agreements and Disagreements Within and Across Categories of Ratings in Transfer Sessions[a]

Frequency of Agreements Within Categories		Frequency of Disagreements Across Categories	
Independence		**Pseudoindependence--Independence**	
ghly--Highly	0	Highly pseudo.--Highly inde.	0
derately--Moderately	10	Moderately pseudo.--Highly inde.	0
derately--Highly	8	Highly pseudo.--Moderately inde.	0
n	18	Moderately pseudo.--Moderately inde.	0
		n	0
Pseudoindependence		**Independence--Dependence**	
ghly--Highly	0	Highly inde.--Highly depend.	1
derately--Moderately	0	Moderately inde.--Moderately depend.	1
derately--Highly	0	Highly inde.--Moderately depend.	4
n	0	Moderately inde.--Moderately depend.	7
		n	13
Dependence		**Pseudoindependence--Dependence**	
ghly--Highly	1	Highly pseudo.--Highly depend.	0
derately--Moderately	8	Moderately pseudo.--Highly depend.	1
derately--Highly	4	Highly pseudo.--Moderately depend.	0
n	13	Moderately pseudo.--Moderately depend.	1
		n	2
N	31	N	15

[a] See footnote, Table 14.

Table 16

Checklist Designations: Global Ratings

vs. Z-Score Classifications[a]

N=29[b]

Global Rating	Z-Score Classification	
	Independent	Non-Independent
Independent	13	8
Dependent	0	1
Pseudoindependent	0	7
Totals	13	16

[a] Data analysis: Chi-square = 8.976; two degrees of freedom; $p < .025$; contingency coefficient = .486.

[b] N of observed Ss in the classrooms = 40. Eliminated from this analysis were: 9 Ss (see Table 11) whose status was designated mixed on the basis of Z-score profiles (note, one of these Ss was also a "can't rate" as to globally rated status); and 2 Ss who represented disagreements as to category between the global checklist raters.

Table 17

Global Checklist Ratings vs. Behavioral Transfer Session Ratings[a]

N=22[b]

Behavioral Transfer Session Rating	Global Checklist Rating	
	Independent	Non-Independent[c]
Independent	8	3
Non-Independent	11	0
Totals	19	3

a
Data analysis: Chi-square=1.544; one degree of freedom; p value is non-significant; phi coefficient =.27. Chi-square value is Yates-corrected.

b
N of observed Ss in the classrooms = 40; one S was absent from the transfer session so that all analyses for experimental Ss in the transfer sessions are based on an N of 39. Of these Ss, for purposes of the above analysis, the following Ss were eliminated: 14 Ss representing disagreements between raters in the behavioral sessions, and the three Ss representing disagreements or "can't rate" of the global raters of the checklist (interestingly, 2 of these 3 Ss also represented disagreements between raters in the transfer sessions).

c
This category represents a "collapsed" category of Dependent and Pseudoindependent global ratings.

Table 18

Z-Score Checklist Designations vs. Ratings in the Transfer Situation[a]

N=18[b]

Ratings in Transfer Sessions	Z-Score Checklist Designation	
	Independent	Non-Independent
Independent	4	4
Dependent	6	3
Pseudoindependent	1	0
Totals	11	7

a
Data analysis: Chi-square = 1.168; two degrees of freedom; p value is nonsignificant; contingency coefficient = .247.

b
N of observed Ss in the classroom = 40; one S was absent from the transfer session so that all analyses for experimental Ss in the transfer sessions are based on an N of 39. Of these Ss, for purposes of the above analysis, the following Ss were eliminated: 9 Ss for whom the Checklist Z-score designations were "mixed" (four of these were also disagreements between the raters in the behavioral sessions); and 12 Ss representing disagreements between raters in the behavioral sessions.

Table 19

Frequency Distributions of Global Checklist Ratings, Z-Score Checklist Designations, and Behavioral Sessions Ratings (Experimental Subjects Only) by Length of Exposure to Institute Program ("Filler Status")

Subject Classification	N[a]	Length of Exposure to Institute Program[b]				
(1)		E	FK	F1	F2	F3
Global Checklist Rating	37					
Independent		11	1	7	2	8
Dependent		0	0	0	1	0
Pseudoindependent		4	1	1	0	1
(2)						
Z-Score Checklist Designation	31					
Independent		4	1	4	1	3
Non-Independent		10	1	4	1	2
(3)						
Behavioral Session Rating	23					
Independent		3	0	4	1	4
Dependent		4	1	2	1	2
Pseudoindependent		0	0	0	0	1

[a] See Footnote (a) Table 20 for explanation of these Ns.

[b] Groups designated as E entered the program at the pre-kindergarten level in September, 1964.
Groups designated as FK entered the program at the kindergarten level in September, 1965.
Groups designated as F1 entered the program at the first grade level in September, 1966.
Groups designated as F2 entered the program at the second grade level in September, 1967
Groups designated as F3 entered the program at the third grade level in September, 1968.

Table 20

Checklist and Transfer Session Ratings of Experimental Subjects and Length of Exposure to the Institute's Program ("Filler" Status)

	N[a]	Comparison Groups[b]	Chi-Square Value	Degrees of Freedom	p value
(1)	37	Global Checklist Ratings X F3 Ss vs. all others (3x2 table)	.872	2	ns
(2)	31	Z-score Checklist Designations X F3 Ss vs. all others (2x2 table)	.799[c]	1	ns
(3)	23	Behavioral (Transfer) Session Ratings X F3 Ss vs. all others (3x2 table)	2.847	2	ns

a
N=40 (classroom observations). For comparison (1), N=37, since 3 Ss represented disagreements or can't rate in re: the global ratings. For comparison (2), N=31 since there were 9 "mixed" Z-score classifications. For comparison (3) N=23, since there were 39 experimental Ss in the transfer session and 16 represented disagreements between the raters.

b
See Table 19 for explanations of "filler" status. For example, F3 refers to all experimental subjects who entered the Institute's program in 1968.

c
Yates-corrected (2x2 table).

Table 21

Chi-Square Comparisons of Ratings of Behavioral Sessions:

Experimental vs. Control Subjects for Both Schools

Combined and for Schools Held Separately[a]

Comparison	Rating	Exper.	Control	N[b]	Chi-Square	p value
		Both Schools Combined				
(1)						
With "low confidence" ratings included	Indep.	7	13		.863	ns
	Depend.	8	6			
N		15	19	34		
(2)						
With "low confidence" ratings excluded (even if one rater used this category)	Indep.	4	9		.327	ns
	Depend.	6	6			
N		10	15	25		
(3)	Public School 175 (With "low confidence" ratings included)					
	Indep.	3	5		.059	ns
	Depend.	4	3			
N		7	8	15		
(4)	Public School 90 (With "low confidence" ratings included)					
	Indep.	4	8		.283	ns
	Depend.	4	3			
N		8	11	19		

[a] All comparisons are based on four-fold tables (one degree of freedom) with chi-square values that are Yates-corrected. The experimental group contains only those Ss who have had one or more years of exposure to the Institute program.

[b] See text, Part B, end of Chapter 3, for explanation of these Ns.

Geometric Design (A) for Task (2)

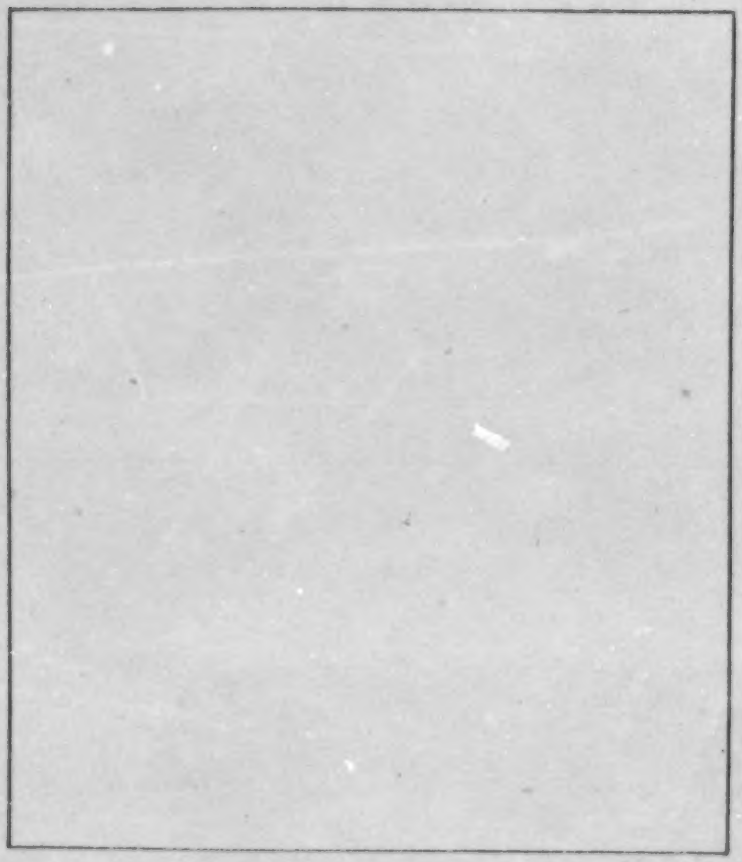


Figure 1

Solution for Design (A)--Task (2)

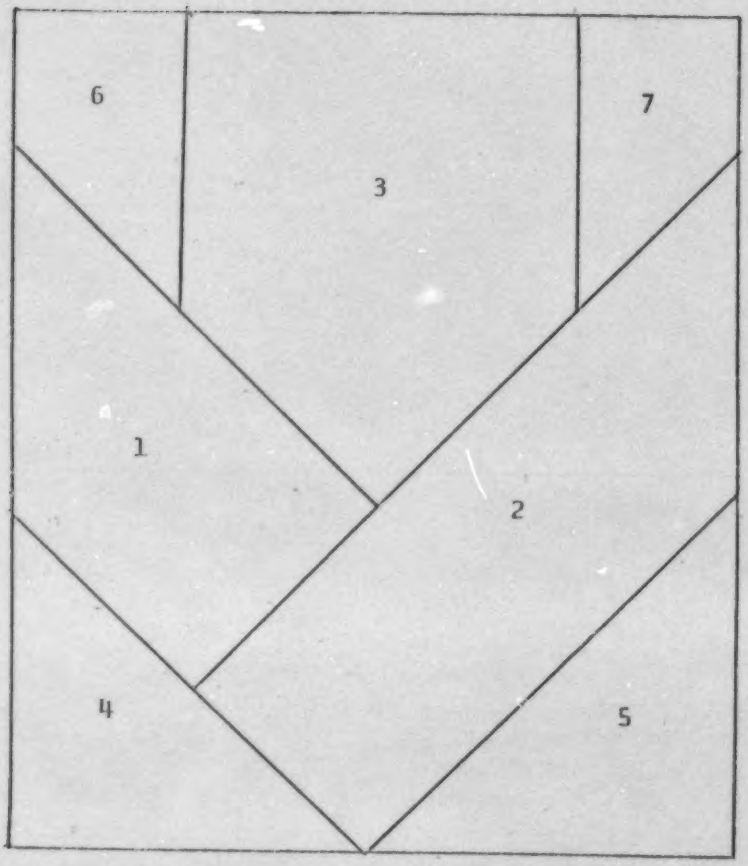


Figure 2

Geometric Design (B) for Task (2)

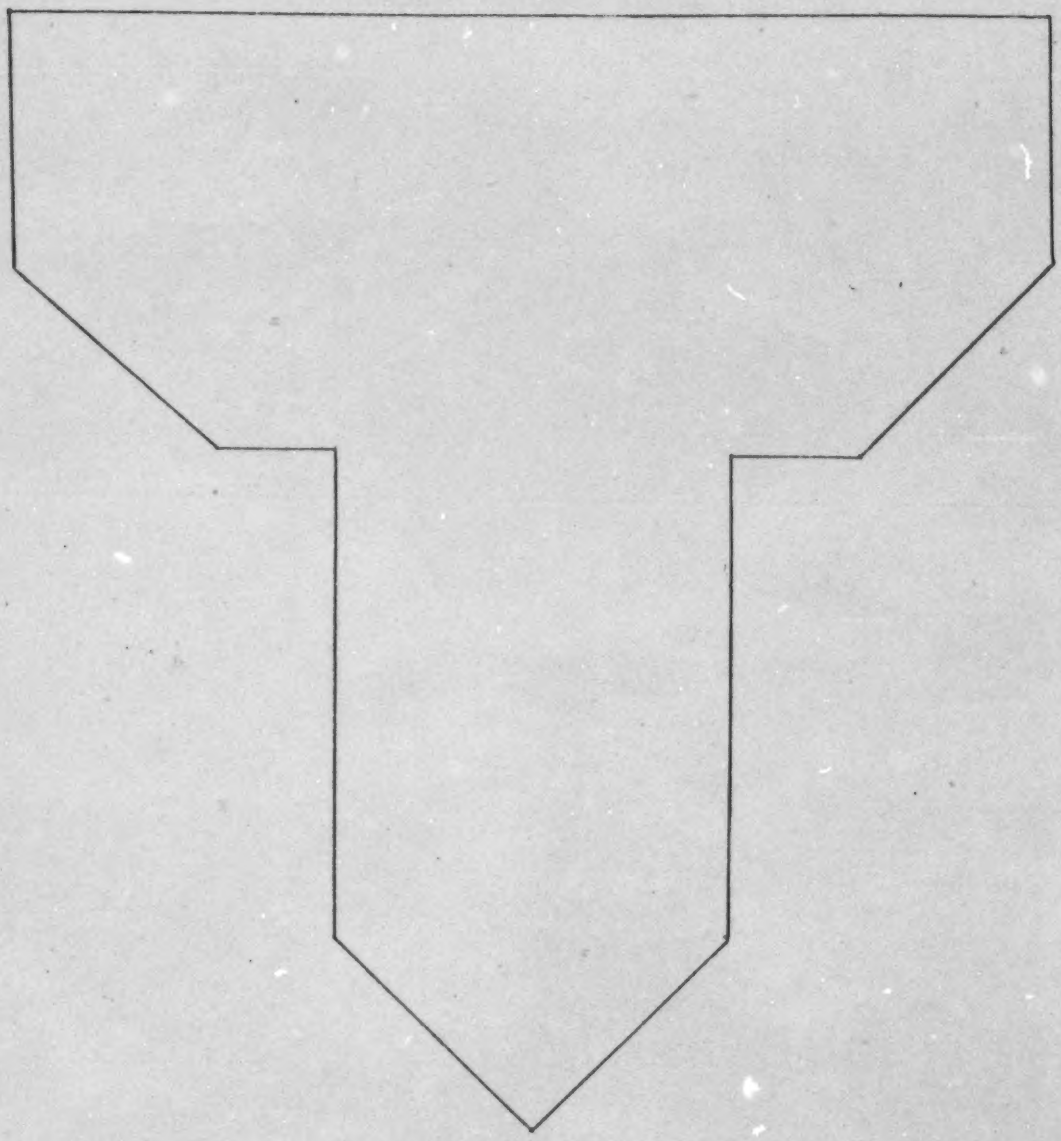


Figure 3

Solution for Design (B)--Task (2)

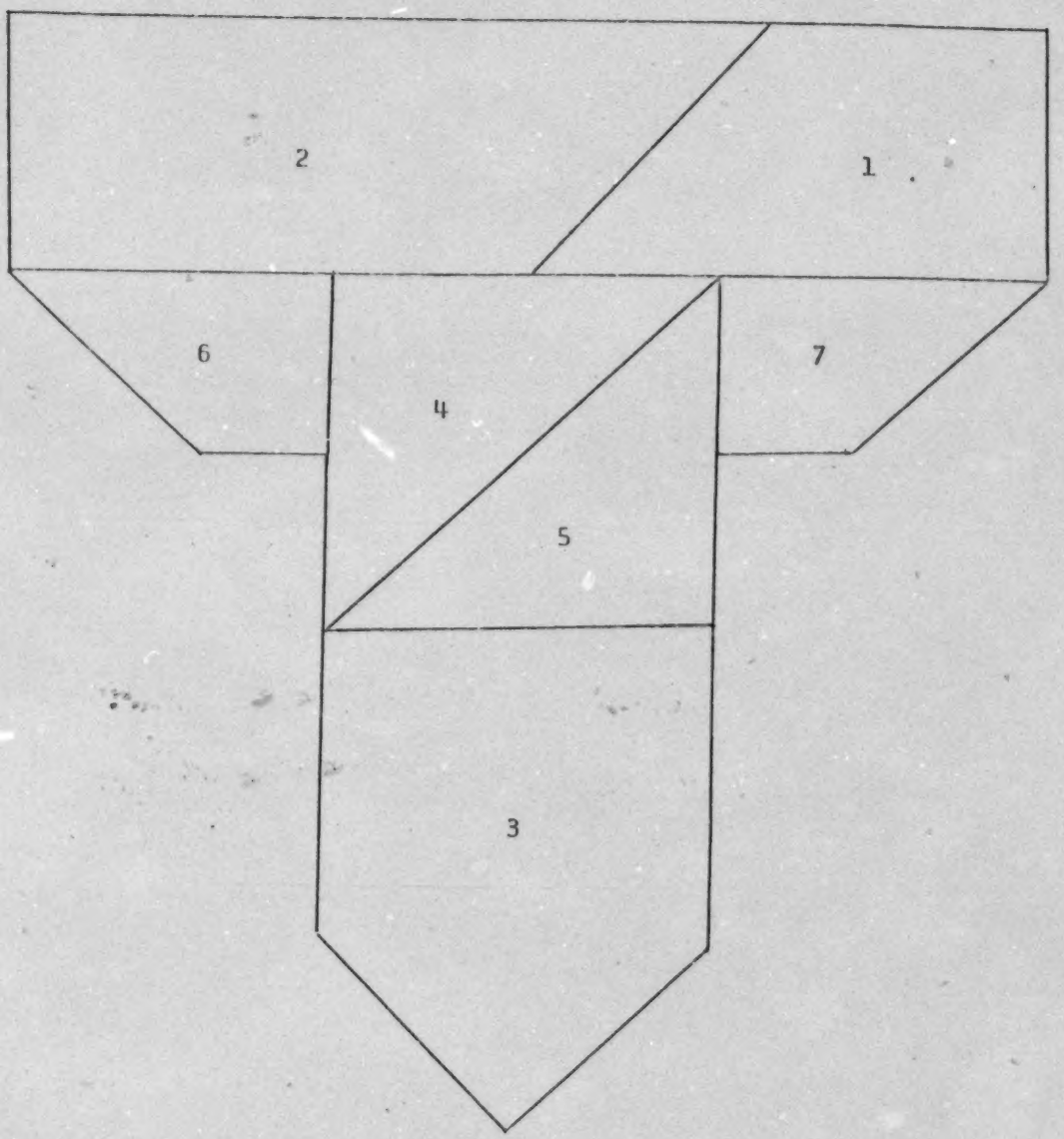


Figure 4

Partial Reproduction of Card in Set (A), Task (3)

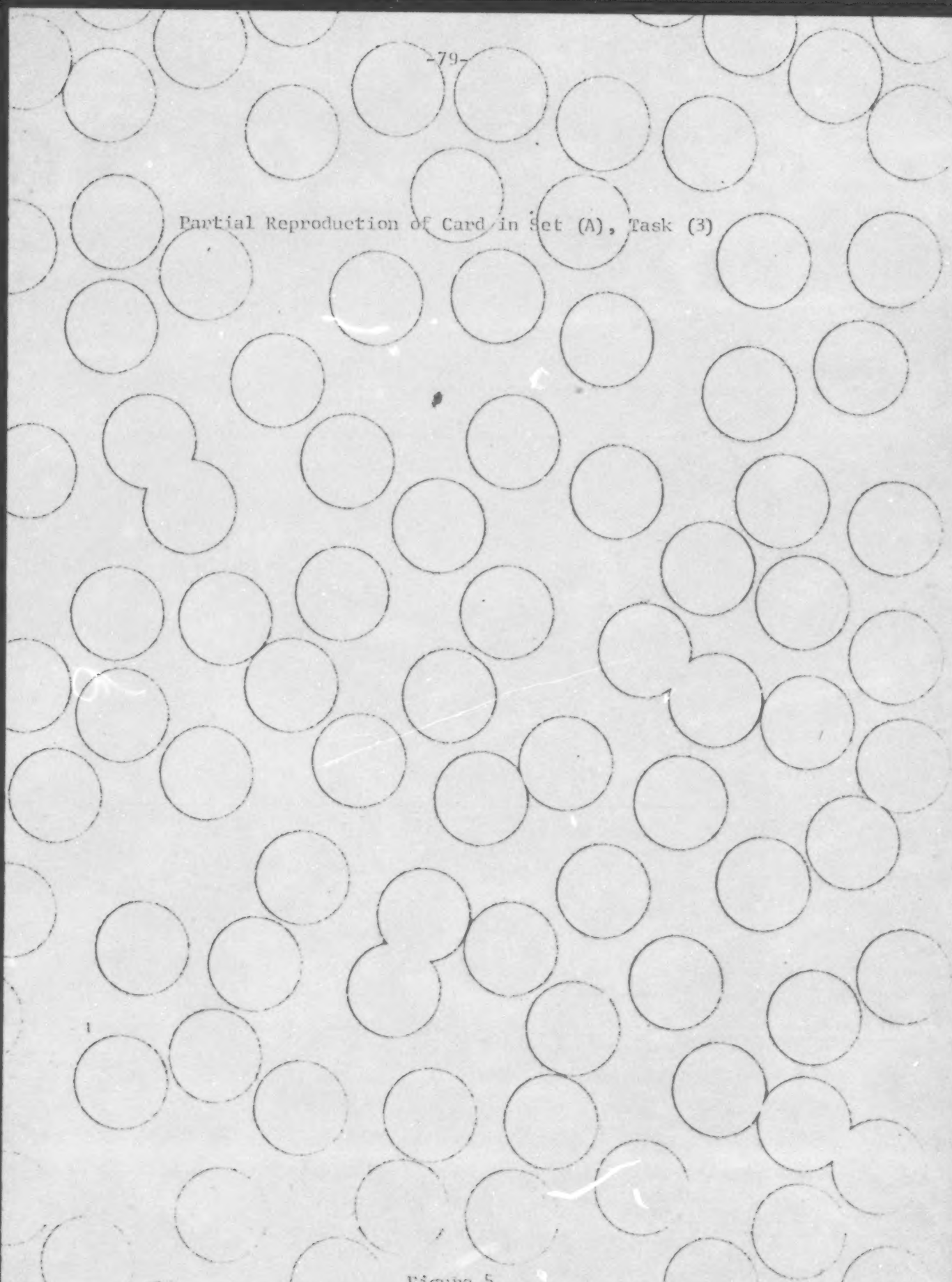


Figure 5

Partial Reproduction of Card in Set (B), Task (3)

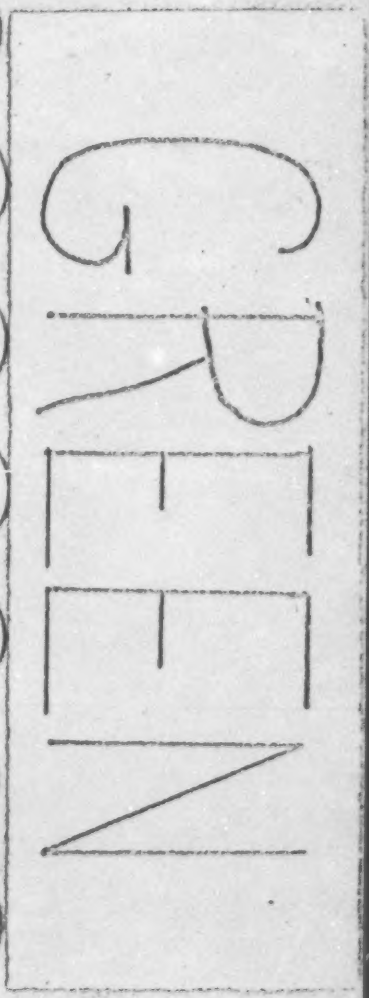


Figure 6

Partial Reproduction of Card in Set (C), Task (3)

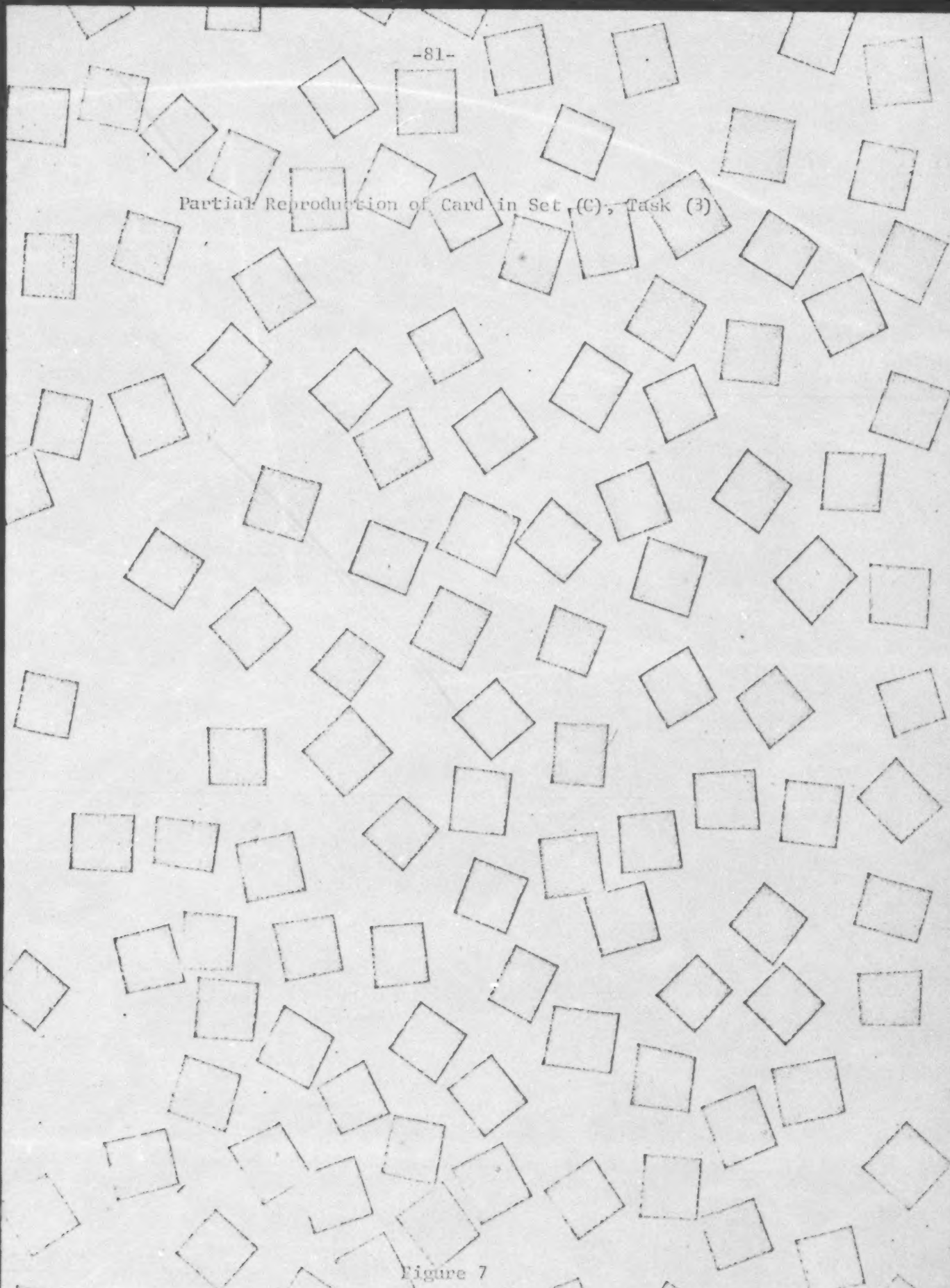
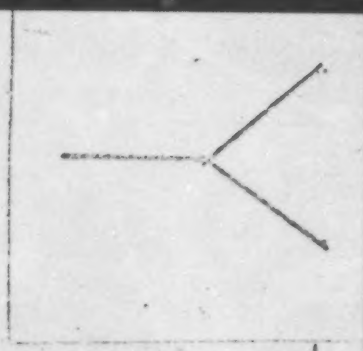


Figure 7



Partial Reproduction of Card in Set (D), Task (3)

Figure 8

Cup Stimulus, Set (E), Task (3)

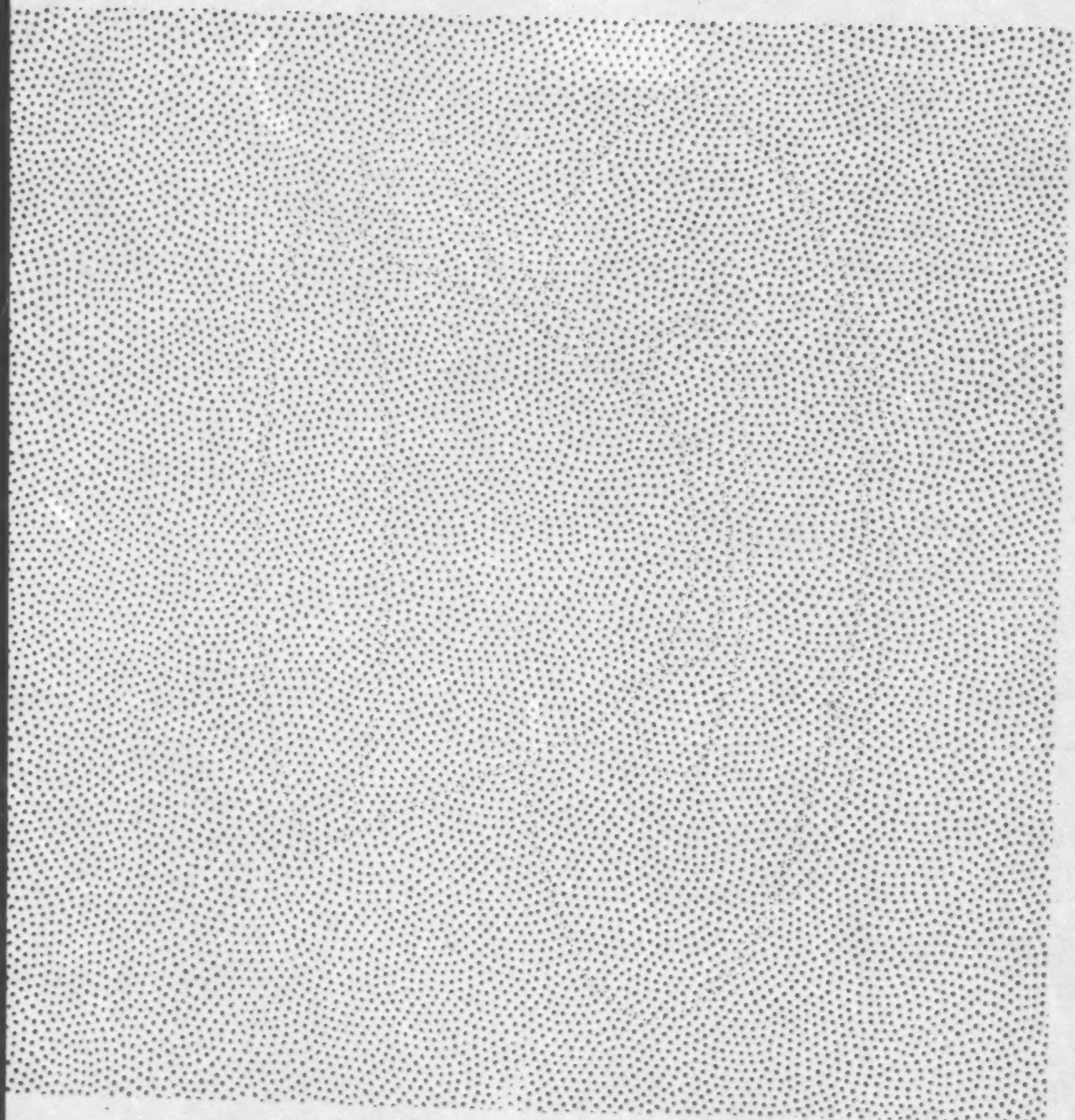


Figure 9

Horse Stimulus, Set (F), Task (3)

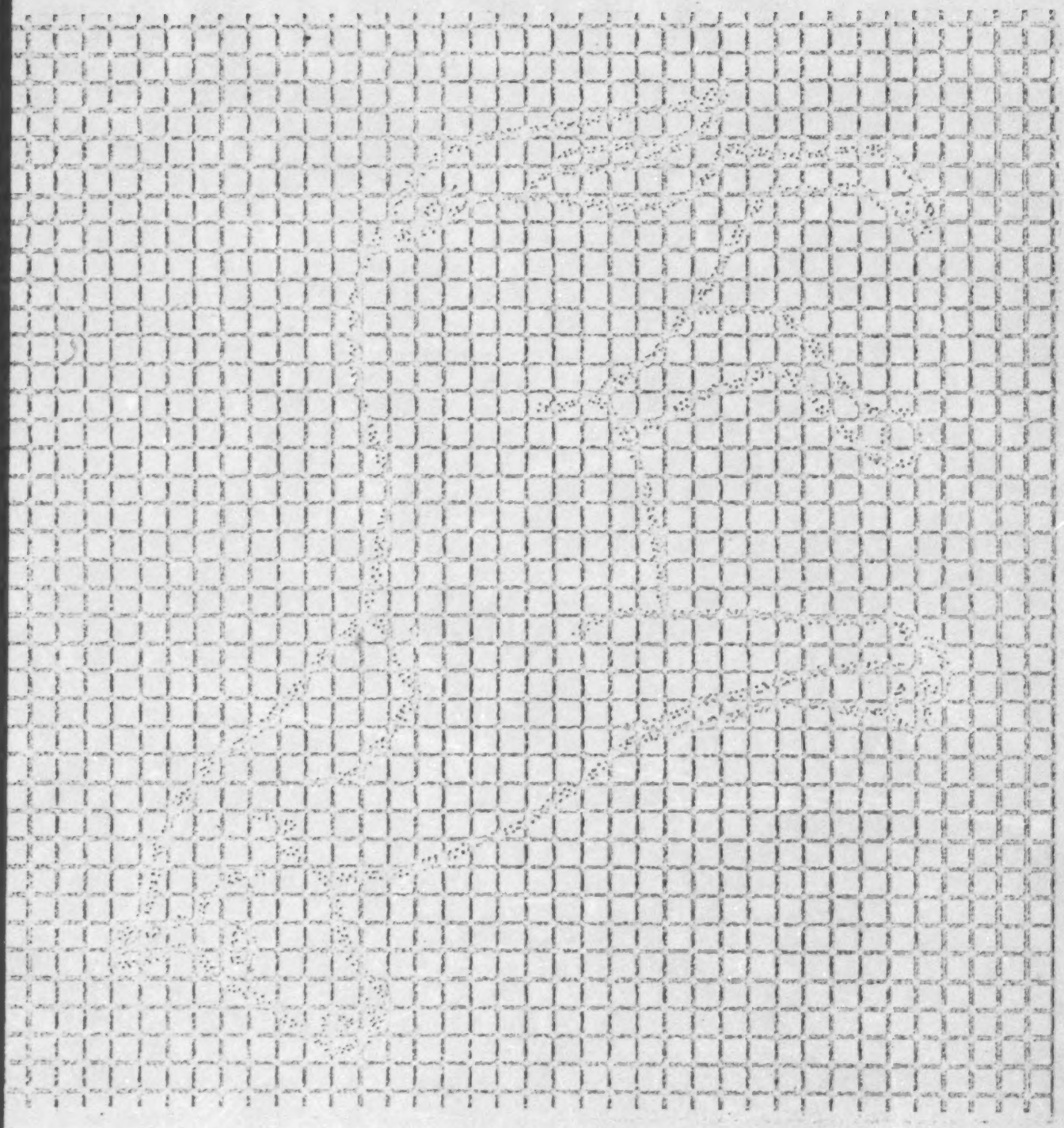


Figure 10

A Page from Booklet, Task (4)

252

Figure 11

Doll-Like Figure used for Demonstration Purposes, Task (5)

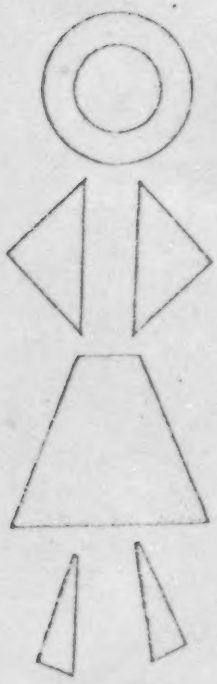


Figure 12

Appendix A

The Classroom Behavior Checklist

CODE _____ SEX _____ DATE _____

OBSERVER _____ TIME STARTED _____

OBSERVATION NUMBER _____

SCHOOL _____ TEACHER _____

		Frequently	Moderately Often	Rarely Often	Never	No Opportunity to Observe; Can't Rate
(1)	withdraws in the face of difficulty					
(2)	is inattentive; is easily distracted by things going on around him					
(3)	uses materials and equipment in a rough or destructive manner					
(4)	volunteers contributions to class discussions and projects					
(5)	helps, sympathizes, shows consideration of others; is thoughtful; offers praise, guidance, assurance					
(6)	sits quietly and attentively when task requires					
(7)	brags, shows off, displays exaggerated opinion of own abilities					
(8)	seeks attention, i.e., in a manner that seems designed for active notice rather than stemming from a need for help					
(9)	utilizes available free time constructively					
(10)	interferes with the work of others					
(11)	withstands interference while engaged in individual activities, e.g., doing difficult assignment, a puzzle, painting					
(12)	imitates others rather than initiating own activities					
(13)	gets upset when called upon in class; appears fearful, tense, timid					
(14)	takes appropriate initiative in problem-solving situations					
(15)	seeks to be near others					
(16)	does not follow directions					
(17)	requires close assistance or direction in order to work at a task					
(18)	seeks to do things differently from others, even when own method is not effective					
(19)	lets other children impose on him or boss him around					
(20)	tries to figure out things for himself before calling on teacher (leader) or other children for help					

(21) enters readily into new situations; is venturesome, inquisitive, etc., when appropriate to ongoing activities
(22) does not pursue task to completion
(23) seeks constant recognition and/or approval
(24) speaks out of turn; interrupts others; takes other's turn, etc.
(25) attempts to resolve difficulties that arise between himself and other children without appealing to teacher (leader)
(26) approaches new tasks timidly and without confidence; shrinks from trying new things
(27) tends to choose habitual and familiar activities

Appendix B

The Revised Classroom Behavior Checklist

	Frequently	Moderately Often	Rarely	Never	No Opportunity to Observe
(1) withdraws in the face of difficulty					
(2) is inattentive; is easily distracted by things going on around him					
(3) uses materials and equipment in a rough or destructive manner					
(4) helps, sympathizes, shows consideration of others; is thoughtful; offers praise, guidance, assurance					
(5) sits quietly and attentively when task requires					
(6) brags, shows off, displays exaggerated opinion of own abilities					
(7) seeks attention, i.e., in a manner that seems designed for active notice rather than stemming from a need for help					
(8) utilizes available free time constructively					
(9) interferes with the work of others					
(10) withstands interference while engaged in individual activities, e.g., doing difficult assignment, a puzzle, painting					
(11) does not follow directions					
(12) requires close assistance or direction in order to work at a task					
(13) seeks to do things differently from others, even when own method is not effective					
(14) lets other children impose on him or boss him around					
(15) tries to figure out things for himself before calling on teacher (leader) or other children for help					
(16) does not pursue task to completion					
(17) seeks constant recognition and/or approval					
(18) speaks out of turn; interrupts others; takes other's turn, etc.					
(19) attempts to resolve difficulties that arise between himself and other children without appealing to teacher (leader)					

SCHOOL _____ TEACHER _____ OBSERVER _____ DATE _____ SEX _____ GRADE _____ NUMBER _____

Appendix B

Creativity Battery: Scoring

The following scoring procedures were employed for each of the subscales of the creativity battery.

(1) Picture Completion

Fluency. This was scored according to Torrance (1966). Credit was given to each unrepeated figure completed, omitting those items in which the lines were not used.

Elaboration. This was scored primarily according to the Torrance format. Credit (one point for each) was given to each elaborative detail of the total response. Credit (one point each) was also given for color, deliberate shading, decoration, each major variation of design which was meaningful with reference to the total response, and each elaboration in the title beyond the minimum descriptive label. No credit was given for essential parts of a picture (e.g., for human and animal faces, eyes, nose, and mouth were considered essential parts). In addition, the following considerations were made, over and beyond those of Torrance. First, for an item to be scored, the basic stimulus item had to be modified in some way. Second, only the main figure of the S's drawing was considered to be the figure made by the S even if his title referred to another particular aspect of the picture.

Flexibility. The flexibility score was obtained by counting the number of different categories into which the responses fell. These categories are reported in Torrance (1966). Here again, it should be noted that the figure drawn, rather than the title, was used in determining category; if two unrelated objects appeared in the same picture,

a flexibility category was designated for each. "Don't know" responses or arbitrary titles for a drawn figure were categorized as "designs" in this system.

Originality. A score of zero or one point was given for each item. If, on the basis of the frequency distribution reported by Torrance (1966) or that resulting from the current sample, two or more children drew an object, the score for that drawing was zero (0). This procedure allowed some flexibility with this sample which was unlike that used in the original standardization procedure. Zero score items are given below:

Item 1: bird, cloud, eye glasses, face, flower, monster, spaceman, valentine

Item 2: campfire, clothesline, design, figure, flower, house, K, rocket (jet), slingshot, tree, branch

Item 3: bananas, bow and arrows, chair, eyes, figure, monster, slide

Item 4: animal, bird, candle, design, duck, ear, fish, whale, glasses, girl, lady, hair, wig, P, smoke, snail, waves, worm

Item 5: banana, boat, bow and arrow, circle, dish, cup, egg, face, hammock, teeth

Item 6: chair, face, figure, fish, kite, lightning, mountains, pot, steps, tree, Z

Item 7: car, truck, key, pipe, scale, spoon, ladle, question mark

Item 8: bib, dress, figure, rocket, shield, tree

Item 9: animal ears, faces, figure, fish, M, mountain, hill, avalanche

Item 10: animal, bird, duck, hill, human face, tree, tree house, witch

One point was given to an item if that item appeared only once in the distributions referred to above. But to receive a score of one, the item had to posses "creative strength." To possess creative strength, the response had to indicate that the stimulus lines were used in a relevant, constructive way, though the picture need not have been drawn well. Furthermore, the picture drawn had to reflect the title used by the child.

(2) Parallel Lines

Fluency. Credit was given to all unrepeated figures which were completed on this subscale. According to Torrance's system, this tally reflected only those items which were relevant. That is, for all items, both lines of the stimulus had to be part of one picture. If not, lines had to be related by category, function, or sameness. "Don't know" responses were considered irrelevant. For this subscale, repetitions were defined as those responses with the same titles (even if drawn differently), and those pictures which elicited the same figural response (even if there was a positional variation). Finally, if the same figural elements were present, although magnified, reduced or elaborated, the item was considered a repetition of an earlier one. In the case of doubt concerning whether the response was a repetition, the title was taken into account. If the title placed the picture into the same category or genre as an earlier picture, the response in question was considered a repetition.

Elaboration. The same criteria and scoring system used for Picture Completion was used for Parallel Lines. If an S made only a slight change from a preceding object in order to draw a "new" object, credit was not given for elaboration.

Flexibility. As with the scoring for the Picture Completion Test, Torrance's categories were used in deriving scores for this variable. Again, the figure drawn, rather than the title, was used in determining category. "Don't know" responses and arbitrary names for pictures were considered designs.

Originality. The same procedures established for the Picture Completion Test were followed here. That is, distributions of responses based on the standardization population as well as the current sample were employed in scoring the originality of parallel lines. The following zero score items resulted for this subscale. If a response did not appear in the combined distribution it was given a score of one point. Because the stimulus elements of all items are the same, zero score response for all items come from the same pool, as listed below:

airplane, aquarium

bag, baggage, ball, barn, bat, bed, blocks, blackboard, boat, bomb, book, bow and arrow, bow, box, brick, bridge, bucket, brush, butterfly

cabinet, cage, cake, calendar, can, candle, candy, car, chimney, clock, crosses, cup, crayon

design, door, dresser

elevator, envelope

faces, firehouse, flags, flower

garbage can, gift, glass, goal post, gun

hat, highway, hourglass, house

ice cream cone

jack-in-the-box, jail

kite

ladder, leg, letters, lines, lollipops

missile, monster

package, pail, paintings, pants, paper, pencils, persons, photograph, pole, prison

radio, rectangle, ribbon, robot, rocket

school, shelf, shirt, signal, sign, silo, slide, space ship, square, star, stilts, store, swing, sword

t.v., table, tent, tic-tac-toe, tie, tracks, train, tree, triangle, tower

window

(3) Product Improvement

For this subscale, Torrance's scoring procedures were followed fairly closely. For the fluency score, for example, almost any unrepeated changes or additions produced by the S were counted. Responses indicating "making the elephant alive," or repeating an idea, or concretizing or magnifying an idea were not counted. For the flexibility score, Torrance's categories were followed in deriving those categories of responses to be counted. Finally, for the originality score, zero or one point was given for each response.

Zero scores for originality were derived in a manner which maintained equivalence with Torrance's formulations for deriving scores (Torrance, 1966). If a response appeared four or more times in the current distribution of responses and is assigned a weight of one or zero by Torrance, it was given a zero score. If, however, a response appeared three or fewer times in the current distribution and was assigned originality weight of one by Torrance, it was given a one point score. If a response occurred three or fewer times in our sample but

was given no originality weight by Torrance, the item had to show "creative strength" to receive one point. That is, the response had to be relevant and constructive in nature.

The following is a list of obtained responses and their frequencies in the current sample, and the weights assigned by Torrance:

	Current Frequencies	Torrance's Weight
Bigger, make it	4	0
Cage, put it in	2	1
Car, truck (put it in)	1	1
Change to other animal	5	-
Change to other toy	3	-
Circus animal, put it with	1	1
Colors, change	5	0
Clothes, dress or coat, add	2	1
Ears, tied back, down	3	-
Ears, make bigger	1	0
Ears, make flappy	1	1
Eat, make it	5	0
Eyes, open and close	1	-
Eyes, put something pointy in	1	-
Feet, tie together	1	-
Friends, put it with other elephants	1	0
Glasses on it	1	-
Hair, put on	1	-
Interchangeable parts	1	-
Jump, make it	2	0
Jungle, put around him	1	1
Less costly (or more costly) make it	3	1
Nails, put on	2	-
Man, sitting on him, children, etc	1	1
Motion, give it by winding, wheels, motor, etc	10	0
Move in different ways, body	3	0
Name, give it	1	0
Noise, make it, make a	4	0
Other animal, put with	1	1
Position change, e.g., stand it up, bend over, legs straight	3	-
Pull objects, make it (put string on it)	2	1
Real elephant, make it	1	0
Ride it, make it easy to, ride him	3	0
Run, make it	4	0
Seat on back, put	1	1
Sit in a chair	2	-
Sleep, make it	1	1
Smaller, make it (little, skinny)	2	0

Squirt or spray water.....	1	0
Stand on hind legs, front legs, make it.....	1	1
Stuffing, add or subtract.....	4	-
Tail, put down.....	2	-
Tail, make it move.....	3	0
Take it apart.....	5	-
Talk, make it (mechanical).....	3	1
Teeth, put in.....	1	-
Truck, put in.....	2	1
Trunk, make it longer.....	1	0
Trunk, make it shorter.....	1	0
Trunk, blow.....	1	-
Tusks, make big (long, strong).....	3	1
Walk, make it.....	8	0
Write on him.....	1	-

(4) Unusual Uses

Torrance's scoring methods were followed rather closely. For the fluency score, the number of unusual uses mentioned by the child was counted but repetitions were omitted. Thus, if the child suggested different things to make from the box, each object was given credit unless the objects were similar in nature or function. Irrelevancies were also eliminated from scoring procedures. For example, a response such as "put it on the fire escape" was considered irrelevant to the use of the box. Flexibility scores were derived from Torrance's categories of responses for unusual uses in a standardization population. Finally, for originality, a zero or one point score was given to each item. Zero scores were derived in a manner which maintained equivalence with Torrance's formulations for deriving scores. If a response occurred five or more times in the current distribution of responses and was assigned a one or zero originality weight by Torrance, it was given a zero score in our system. If a response occurred four or fewer times in the current distribution of responses and was assigned an originality weight by Torrance, it was given one point in the current scoring procedure. If a response occurred four or fewer times in the

current distribution of responses and was given no originality weight by Torrance, it had to possess "creative strength" to receive an originality score in our procedure. The following is a list of obtained responses and their frequencies in the current sample and the weights for originality given to them by Torrance:

	Current Frequencies	Torrance's Weight
Airplane, toy (rocket)	4	0
Animal cage, house, shelter	3	0
Animal, cut out (make)	4	1
Art, drawing or painting on	8	0
Aquarium	1	-
Bed, bed to sleep on	2	0
Basket (Easter)	2	1
Boat	4	0
Book case or shelf	1	0
Book, booklet (make)	4	1
Building, including playhouse, club house, etc.	1	0
Car, racer (toy)	3	0
Carrier, "carry stuff in," carry books, etc. (anything but clothes)	8	0
Chair, something to sit on	8	0
Clothing	1	0
Container for junk, flowers, jewelry, pencils, put anything in (garbage, toys, etc.)	27	0
Decorations (cut out, make designs with)	6	1
Destruction, waste, burn, throw away	11	0
Doll house	1	0
Doors	3	1
Flower pot	2	1
Games, specified-- punch bag, kick or basketball court	3	-
Games, unspecified (play with)	6	0
Grocery container (food, pack)	6	0
Give away	2	-
Hat	1	0
House, model of, hospital, school	8	0
Make (bank, bathroom, closet, cut-outs, fan, fence, flag, flying saucer, frame, greeting cards, kite, money, newspaper, pencil, stick, window)	1 (for each item mentioned)	-
Make (bookcover, crayon box, geometric shapes /△, □, ○/, numbers and letters /cut-out/, telephone)	2	-
Make (desk)	3	-

Make (clock)	4	-
Make things out of (build things)	2	-
Patch holes	1	-
Pen, animal, babies	2	0
Pets, put in	6	0
Play in (play house, school in)	6	0
Puppets	3	0
Puppet theatre	1	0
Refrigerator	3	1
Sandbox	1	-
Save it	2	-
Science experiments (projects)	2	-
Shield	2	1
Shoes	2	1
Stove, range, etc	1	0
Suitcase (pack clothing)	5	1
Swimming pool	1	-
Target (dart board)	2	1
Train (toy)	2	0
Trap	1	1
Truck (toy)	2	1
Tree	1	-
Wagon, cart, etc. (ride in)	4	1
Write letters on (writing board)	4	0
Wrap it, wrap in it	3	-